METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT

DIRETTORE PROPRIETARIO — DIRECTEUR ET PROPRIÉTAIRE
EDITOR AND PROPRIETOR — HERAUSGEBER UND EIGENTHÜMER

Prof. Dott. Corrado Gini, *Direttore dell'Istituto di Statistica della R. Università di Roma.*

COMITATO DIRETTIVO — COMITÉ DE DIRECTION
EDITORIAL COMMITTEE — DIREKTION-KOMITEE

Prof. F. Bernstein, *früherer Direktor des Instituts für mathematische Statistik der Universität, Göttingen* (Deutschland), *jetzt an der Columbia University* (U. S. A).
Prof. A. E. Bunge, *ya director gen. de Estadística de la Nación, Buenos Aires* (Argentina).
Prof. F. P. Cantelli, *professore di Matematica Attuariale nella R. Università di Roma* (Italia).
Prof. A. Flores de Lemus, *jefe de Estadística del Min. de Hacienda, Madrid* (España).
Prof. M. Greenwood, *professor of Epidemiology and Vital Statistics in the University of London* (England).
Dott. G. Jahn, *directeur du Bureau Central de Statistique de Norvège, Oslo* (Norvège).
Prof. A. Julin, *secrétaire général honoraire du Ministère de l'Industrie, du Travail et de la Prévoyance sociale, Bruxelles* (Belgique).
Prof. H. W. Methorst, *directeur de l'Office permanent de l'Institut International de Statistique et du Bureau central de Statistique, La Haye* (Pays-Bas).
Prof. W. F. Ogburn, *professor of Sociology in the University of Chicago* (U. S. A.).
Prof. R. Pearl, *director of the Department of Biology of the School of Hygiene and Public Health, Baltimore* (U. S. A.).
Prof. H. Westergaard, *professor in the University of Copenhagen* (Denmark).

AMMINISTRATORE — ADMINISTRATEUR — MANAGER — VERWALTER
Dott. Silvio Orlandi, *Istituto di Statistica della R. Università di Roma.*

SEGRETARI DI REDAZIONE — SECRÉTAIRES DE RÉDACTION
EDITORIAL SECRETARIES — REDACTIONSSECRETAERE
Prof. Luigi Galvani — Prof. Mario Saibante

Vol. XII - N. 3 **5-XII-1935.**

SOMMARIO — SOMMAIRE — CONTENTS — INHALT

C. Gini. *Su la determinazione dei quozienti di eliminazione e in particolare sui metodi delle durate esatte e delle durate medie nell'ipotesi di saggi istantanei di eliminazione costanti* . . . Pag. 3
S. Kullback. *A note on the distribution of a certain partial belonging coefficient* . . . . . » 65
A. Linder. *« Wahrscheinlichkeitsansteckung » und Differenzengleichungen* . . . . . » 71
E. Zwinggi. *Zur Frage des Bearrungszustandes* . . . . . » 91
V. Castellano. *Recente letteratura sugli indici di variabilità* . . . » 101

ROMA
AMMINISTRAZIONE DEL «METRON»
R. UNIVERSITÀ — ISTITUTO DI STATISTICA

ARTICOLI GIUNTI ALLA RIVISTA CHE VERRANNO PUBBLICATI NEI PROSSIMI NUMERI.

(*Secondo l'ordine d'arrivo*).

ARTIKEL DIE AN DIE ZEITSCHRIFT ANGELANGT SIND UND WELCHE IN DEN NACHFOLGENDEN NUMMERN ERSCHEINEN WERDEN.

(*Nach der Reihenfolge des Eingangs*)

ARTICLES REÇUS PAR LA REVUE ET À PARAÎTRE PROCHAINEMENT.

(*D'après la date de reception*).

ARTICLES RECEIVED BY THE REVIEW WHICH WILL BE PUBLISHED IN FUTURE ISSUES.

(*According to date of receipt*)

S. Koller — *Die Analyse der Abhängigkeitsverhältnisse in zwei Korrelationssystemen.*

S. Kullback — *A Note on the multiple Correlation Coefficient.*

G. Pietra — Necrologie: *March, Andréadès, Czuber, Charlier Knibbs.*

Gli Autori degli articoli inviati per la pubblicazione nella Rivista, rinunciano in favore della medesima alla proprietà letteraria degli articoli stessi, qualora vengano pubblicati.

Les Auteurs des articles envoyés à la Revue, pour y être publiés, renoncent, en faveur de celle-ci, à la propriéte littéraire de leurs articles, s'ils sont acceptés.

The Authors of papers sent for publication in the Review are supposed to give up their copyright in favour of the Review if the papers are published.

Die Verfasser der zur Veröffentlichung in der Zeitschrift zugesandten Aufsätze, werden, falls selbige veröffentlicht werden, auf ihre Verfasserrechte zu Gunsten der Zeitschrift verzichten müssen.

CORRADO GINI

Su la determinazione dei quozienti di eliminazione e in particolare sui metodi delle durate esatte e delle durate medie nell'ipotesi di saggi istantanei di eliminazione costanti

Si abbia in osservazione un gruppo di individui esposti al rischio di un dato evento, il cui verificarsi elimina il colpito dal gruppo e sia $i \,|\!-\, i + k$ il periodo di calendario durante il quale noi sottoponiamo il gruppo alla nostra osservazione, allo scopo di determinare la frequenza con cui i componenti vengono eliminati dall'evento in questione durante l'intervallo unitario $t \,|\!-\, t + 1$ della loro esistenza. In quanto a t esso potrà rappresentare il tempo che, all'inizio dell'intervallo, sarà trascorso a partire dalla nascita e costituirà allora l'età dell'individuo, o a partire dalla conclusione di un contratto di assicurazione, e esprimerà allora la durata dell'assicurazione, o a partire da altro avvenimento qualsiasi, per esempio dalla pubertà o dal matrimonio, e via dicendo.

Tra i *quozienti di eliminazione* che misurano detta frequenza vi ha luogo a distinguere :

A) La *probabilità di eliminazione* q_t , che si ottiene ragguagliando il numero degli individui eliminati d_t durante l'intervallo $t \,|\!-\, t + 1$ al numero degli individui eliminabili ε_t durante detto intervallo. Perchè possa parlarsi di probabilità di eliminazione, conviene che gli ε_t individui eliminabili durante detto intervallo siano anche ugualmente eliminabili, nel senso che le condizioni di osservazione non permettano di attribuire alla eliminazione di alcuni individui una probabilità maggiore che alla eliminazione di altri. Non sempre la determinazione di ε_t riesce facile. Quando le condizioni di osservazione sono tali che i componenti il gruppo vengono seguiti per tutto l'intervallo $t \,|\!-\, t + 1$, il numero degli individui eliminabili ε_t coincide col numero E_t degli esposti al

rischio al principio dell'intervallo, salvo in casi particolari di cause di eliminazione temporanee, di cui diremo a suo luogo (cfr. pag. 45).

B) Il *saggio di eliminazione* S_t, che si ottiene ragguagliando il numero degli eliminati d_t al numero medio M_t degli esposti al rischio durante l'intervallo in parola.

Per esempio, il numero dei morti da 0 ad 1 anno, ragguagliato al rispettivo numero dei nati vivi, dà la probabilità di morte durante il primo anno di vita ; ragguagliato al rispettivo numero medio dei viventi da 0 ad 1 anno, dà il saggio di mortalità nel primo anno di vita.

Se chiamiamo *intervallo istantaneo* o *istante* x una frazione piccolissima $x \,|\!-\, x + dx$ dell'intervallo unitario $t \,|\!-\, t + 1$, il rapporto del numero degli eliminati al numero degli esposti al rischio nell'istante x si indica con $\mu_x\, dx$, e μ_x si chiama *saggio istantaneo di eliminazione* o *forza di eliminazione nell'istante* x (1).

S_t è la media ponderata degli n valori di μ_x in cui ad ogni μ_x viene attribuito un peso proporzionale al rispettivo numero E_x degli esposti al rischio nell'istante x.

* * *

Consideriamo dapprima il caso in cui le condizioni di osservazione sono tali che tutti i componenti il gruppo possano essere osservati durante tutto l'intervallo $t \,|\!-\, t + 1$. Ed esaminiamo le relazioni che, in tal caso, passano tra i valori di E_t e M_t corrispondenti a varie ipotesi. Ciò è utile, non solo dal punto di vista teorico, ma anche dal punto di vista pratico in quanto talvolta si conosce il numero E_t degli esposti al rischio all'inizio dell'intervallo $t \,|\!-\, t + 1$; tale altra, invece, il numero medio M_t degli

---

(1) L'espressione saggio istantaneo di eliminazione può però indurre in equivoco, in quanto può far credere che il valore di detto saggio dipenda solo dai numeri degli eliminati e degli esposti al rischio nell'istante x, mentre esso dipende anche dalla lunghezza dell'intervallo unitario $t \,|\!-\, t + 1$; in realtà, il così detto saggio istantaneo di eliminazione nell'istante x fornisce « il valore del saggio di eliminazione che si otterrebbe in un intervallo unitario $t \,|\!-\, t + 1$, qualora persistesse, per tutti gli istanti dell'intervallo, il rapporto tra eliminati e esposti al rischio verificatosi nell'istante x ». Secondo che l'intervallo è di un mese, di un anno, di un decennio, potrebbe propriamente parlarsi di *saggio mensile, annuale, decennale di eliminazione, corrispondente alla forza di eliminazione nell'istante* x.

esposti al rischio durante il detto intervallo. È ovvio che tali relazioni dipendono dalla maggiore altezza del saggio di eliminazione, del pari che dalle sue variazioni durante l'intervallo $t \mid\!— t+1$.

Per ciò che concerne i quozienti di mortalità, i quali sono, tra i vari quozienti di eliminazione, i più studiati, si ammette comunemente che le relazioni tra E_t e M_t siano espresse, salvo per i primi anni di vita, dalla formula di Becker

$$E_t = M_t + \frac{d_t}{2}$$

da cui

$$M_t = E_t - \frac{d_t}{2} = E_t \frac{1}{1 + \frac{S_t}{2}} \qquad (1)$$

Tali formule corrispondono all'ipotesi, che chiameremo ipotesi *a*), che il numero dei morti sia uguale nei successivi istanti dell'intervallo $t \mid\!— t+1$, ciò che equivale ad ammettere che il saggio istantaneo di mortalità vada crescendo in progressione armonica dal principio alla fine del periodo di osservazione.

Sembra che valga la pena di ricercare se non sia molte volte consigliabile di far ricorso ad ipotesi diverse.

Una ipotesi che in molti casi si avvicina meglio alla realtà è certo quella di supporre (ipotesi *b*) che, nei successivi istanti dell'intervallo $t \mid\!— t+1$, rimanga costante, non il numero degli eliminati, ma il saggio istantaneo di eliminazione. E tale ipotesi appare giustificata, in base al principio di indifferenza, quando non si conosca nulla circa le variazioni di detto saggio durante l'intervallo considerato.

Vediamo dunque quali sieno le relazioni tra E_t ed M_t nell'ipotesi *b*), che il saggio istantaneo di eliminazione resti costante negli n intervalli istantanei, in cui si può dividere l'intervallo unitario $t \mid\!— t+1$, vale a dire nell'ipotesi che sia $S_t = \mu_x$, essendo $t \leqslant x < t+1$.

Il numero n di detti intervalli istantanei, od istanti, essendo sufficientemente grande da poter considerare, senza errore sensibile, costante, in ciascuno di essi, il numero degli esposti al rischio, se E_t è il numero degli esposti al rischio nel primo istante,

$$E_t - E_t \frac{S_t}{n} = E_t \left(1 - \frac{S_t}{n}\right)$$

sarà il numero degli esposti al rischio nel secondo istante,

$$E_t\left(1-\frac{S_t}{n}\right)-E_t\left(1-\frac{S_t}{n}\right)\frac{S_t}{n}=E_t\left(1-\frac{S_t}{n}\right)^2$$

sarà il numero degli esposti al rischio nel terzo istante,

$$E_t\left(1-\frac{S_t}{n}\right)^{n-1}$$

sarà il numero degli esposti al rischio nell'n^{mo} istante.

Il numero medio degli esposti al rischio negli n istanti può scriversi

$$M_t=E_t\frac{1}{n}\sum_{x=0}^{n-1}\left(1-\frac{S_t}{n}\right)^x$$

o anche, per la nota formula che dà la somma dei termini di una progressione geometrica :

$$M_t=E_t\frac{1}{n}\frac{1-\left(1-\frac{S_t}{n}\right)^n}{\frac{S_t}{n}}=E_t\frac{1-\left(1-\frac{S_t}{n}\right)^n}{S_t}$$

da cui approssimativamente, per essere n grandissimo,

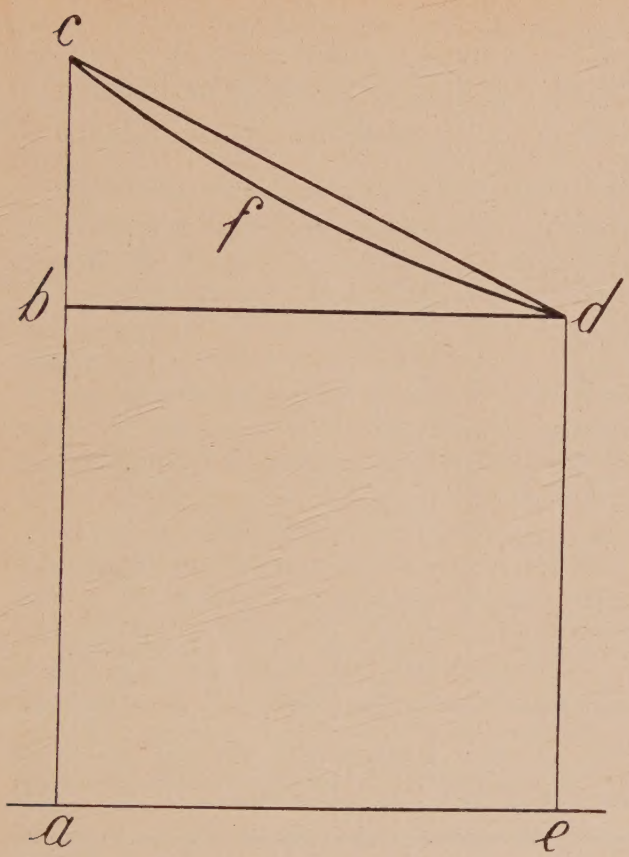
$$M_t=E_t\frac{1-e^{-S_t}}{S_t} \qquad (2)$$

È facile intendere che il numero degli esposti al rischio dato dalla (2), in base all'ipotesi *b*) che il saggio istantaneo di eliminazione sia costante, è minore di quello dato dalla (1), in base all'ipotesi *a*) di un numero costante di eliminati nei successivi istanti, ed è maggiore di quello che si otterrebbe (ipotesi *c*)) se tutta la eliminazione avvenisse nel primo istante, nel qual caso sarebbe :

$$M_t=E_t-d_t=E_t\frac{1}{1+S_t} \qquad (3)$$

Nella figura seguente, il segmento $\overline{a\,c}$ indica il numero degli esposti al rischio all'istante t, il seguento $\overline{e\,d}$ il numero dei sopravviventi all'istante $t+1$, la superficie del trapezio $a\,c\,d\,e$ dà il

numero totale degli esposti durante l'intervallo $t \mid\!- t+1$ nell'ipotesi (*a*), la superficie del rettangolo *a b d e* il numero totale degli esposti nell'ipotesi (*c*), la superficie *a c f d e* il numero totale degli esposti nell'ipotesi (*b*).



Dalle (1), (2) e (3) si ricavano immediatamente (1) le rispettive relazioni tra la probabilità di eliminazione e il saggio di eliminazione:

$$q_t = \frac{S_t}{1 + \frac{S_t}{2}} \quad (1\,a)$$

$$q_t = 1 - e^{-S_t} \quad (2\,a)$$

$$q_t = \frac{S_t}{1 + S_t} \quad (3\,a)$$

Nella tabella I sono esposti, in funzione dei valori di S_t, i valori di q_t in base alle formule (1 *a*), (2 *a*) e (3 *a*), corrispondenti rispettivamente alle ipotesi *a*), *b*) e *c*). I valori calcolati

---

(1) La (2 *a*) si può ricavare anche dalla $q_t = 1 - e^{-\int_t^{t+1} \mu_x\,dx}$ facendo $\mu_x =$ costante e tenendo presente che è allora $\mu_x = S_t$. Questa formula, che esprime la relazione tra la probabilità e il saggio istantaneo di eliminazione, si incontra spesso nella letteratura attuariale. Finora però, che io sappia, non si è pensato a far ricorso all'ipotesi di $\mu_x =$ costante, per quanto questa ipotesi, da una parte, come si è detto nel testo (cfr. pag. 5), appaia in molti casi la preferibile, e, d'altra parte, conduca ad una relazione molto semplice tra la probabilità e il saggio istantaneo di eliminazione.

secondo la formula (2 *a*) restano sempre intermedi tra quelli calcolati secondo le formule (1 *a*) e (3 *a*).

TABELLA I (*).

Valori di S_t	Valori di q_t secondo la formula		
	(1 *a*) $\frac{S_t}{1+\frac{S_t}{2}}$	(2 *a*) $1-e^{-S_t}$	(3 *a*) $\frac{S_t}{1+S_t}$
10	1.66667	0.99995	0.90909
5	1.4286	0.9993	0.8333
2	1.0000	0.8647	0.6667
1.5	0.8571	0.7769	0.6000
1	0.6667	0.6321	0.5000
0.9	0.6207	0.5934	0.4737
0.8	0.5714	0.5507	0.4444
0.7	0.5185	0.5034	0.4118
0.6	0.4615	0.4512	0.3750
0.5	0.4000	0.3935	0.3333
0.4	0.3333	0.3297	0.2857
0.3	0.2609	0.2592	0.2308
0.2	0.1818	0.1813	0.1667
0.1	0.09524	0.09516	0.0909
0.05	0.0487805	0.0487706	0.0476190
0.01	0.00995025	0.00995017	0.00990099
0.005	0.00498753	0.00498752	0.00497512
0.001	0.0009995002	0.00099950017	0.0009990010

(*) I calcoli dei valori esposti in questa tabella sono stati eseguiti dal Dr. Vittorio Castellano, aiuto dall'Istituto di statistica della R. Università di Roma.

Soffermandoci in particolare sulle differenze tra i valori ottenuti secondo le formule (1 *a*) e (2 *a*), si nota che queste sono trascurabili per valori di $S_t \lesssim 0{,}1$, in quanto non concernono che la quinta cifra decimale e la quarta cifra significativa: ancora per valori di $S_t = 0.4$ le differenze non affettano che la terza cifra significativa.

Poichè generalmente i saggi annuali di mortalità al di sopra di due anni di vita non oltrepassano, per la popolazione generale, il 10 %, questi risultati si possono riguardare come la conferma della sufficienza, agli scopi pratici, della formula di Becker.

Nella determinazione di altre probabilità di eliminazione (per esempio, della probabilità annuale di eliminazione per parto

delle coniugate o della probabilità di eliminazione per perdite di guerra dei militari), può invece essere conveniente di ricorrere alla formula da noi proposta. Quando, infine, il valore dei saggi di mortalità è molto elevato, la formula (1 *a*) non si può più usare, in quanto per $S_t > 2$, essa porterebbe a valori della probabilità di eliminazione superiori all'unità.

* * *

Finora si è supposto che le condizioni dell'osservazione sieno tali da poter seguire tutti gli esposti al rischio durante tutto l'intervallo $t \mid\!— t + 1$. In tal caso gli esposti al rischio si possono dire *in osservazione completa*. Ma ciò non avviene quasi mai e talvolta la realtà è così lontana da siffatta ipotesi che dalla divergenza non si può prescindere.

Alle volte gli esposti al rischio entrano in osservazione dopo aver oltrepassato l'istante t della loro esistenza, sia che avessero già oltrepassato tale istante all'inizio i dell'osservazione, sia che siano entrati in osservazione in un momento successivo a i (per esempio, per immigrazione). Viceversa, altre volte gli esposti al rischio escono dall'osservazione prima di aver raggiunto l'istante $t + 1$, sia perchè sopraggiunge prima la chiusura $i + k$ del periodo di osservazione, sia perchè altre cause (per esempio, emigrazione) li eliminano dall'osservazione. In tali casi gli esposti al rischio si diranno *in osservazione incompleta*.

Si presenta, in questi casi, il problema di ricavare da dati, tutti o in parte relativi a esposti al rischio in osservazione incompleta, i quozienti di eliminazione che si sarebbero ottenuti qualora gli esposti fossero stati tutti in osservazione completa, ciò che d'altronde non si può fare senza ricorrere ad ipotesi.

Il procedimento più esatto in uso è quello chiamato appunto *metodo delle durate esatte* (1). Esso porta a determinare un numero virtuale di esposti al rischio in osservazione completa, equivalente, ai fini dell'esposizione al rischio e dell'eliminazione, al numero effettivo di esposti in osservazione incompleta.

---

(1) Esso è dovuto a Thomas G. Ackland. Cfr. Th. G. Ackland : (*I*) *An investigation of some of the Methods for deducing the Rates of Mortality, and of Withdrawal, in Years of Duration, with* (*II*) *the Application of such Methods to the computation of the Rates experienced, and the Special Benefits granted, by Clerks' Associations*, in « Journal of the Institute of Actuaries », Vol. XXXIII, Part I, October 1896, pag. 68, Part II, January 1897, pag. 131.

Si indichi con l_t il numero degli esposti al rischio entrati in osservazione all'istante t ; poichè la durata dell'intervallo $t\,|\!-\,t+1$ è l'unità, l_t indicherebbe anche la durata complessiva della loro esposizione al rischio se questa si prolungasse sino alla fine dell'intervallo $t\,|\!-\,t+1$. Si indichi con u l'istante dell'intervallo $t\,|\!-\,t+1$ in cui l'esposto al rischio entra o esce dall'osservazione, dove è $t < u < t+1$. Infine, si indichi con s_u il numero degli esposti al rischio che esistono all'inizio i dell'osservazione e quindi con $s_u\,(t+1-u)$ la loro durata complessiva di esposizione al rischio, se questa si prolungasse sino alla fine dell'intervallo $t\,|\!-\,t+1$; con r_u il numero degli esposti al rischio che sopravvengono dopo l'inizio i del periodo di osservazione e quindi con $r_u\,(t+1-u)$ la loro durata complessiva di esposizione, se questa si prolungasse sino alla fine dell'intervallo $t\,|\!-\,t+1$, con d_u il numero degli esposti al rischio che restano eliminati dall'osservazione per la causa di eliminazione di cui si vuole determinare la probabilità e quindi con $d_u\,(t+1-u)$ la durata complessiva di esposizione per cui essi restano eliminati dal rischio ; con w_u il numero degli esposti al rischio che restano eliminati dall'osservazione per una causa di eliminazione diversa da quella di cui si vuole determinare la probabilità e quindi con $w_u\,(t+1-u)$ la durata complessiva di esposizione per cui essi restano eliminati dal rischio, con e_u il numero degli individui che rimangono alla chiusura $i+k$ del periodo di osservazione e quindi con $e_u\,(t+1-u)$ la durata complessiva di esposizione per cui essi restano eliminati dal rischio.

Il metodo delle durate esatte si fonda sull'ipotesi (ipotesi d) che n individui in osservazione incompleta, i quali entrano in osservazione nel momento u dell'intervallo unitario $t\,|\!-\,t+1$ e sono seguiti sino alla fine dell'intervallo, equivalgano, ai fini dell'esposizione al rischio e dell'eliminazione, a $(t+1-u)\,n$ individui in osservazione completa.

Il numero virtuale degli esposti in osservazione completa sarà quindi :

$$E_t = l_t + \int_t^{t+1} (s_u + r_u - w_u - e_u)\,(t+1-u)\,d\,u$$

da cui

$$q_t = \frac{d_t}{l_t + \int_t^{t+1} (s_u + r_u - w_u - e_u)\,(t+1-u)\,d\,u} \qquad (5)$$

che dà precisamente la probabilità cercata, determinata col metodo delle durate esatte.

Tale ipotesi si può anche formulare dicendo che la probabilità di eliminazione durante la frazione $u \,|\!-\, t+1$ dell'intervallo $t \,|\!-\, t+1$ è proporzionale all'intervallo $t+1-u$, in modo che si abbia :

$$q_{(u\,|\!-\,t+1)} = (t+1-u)\, q_t \qquad (d)$$

Se ne deduce

$$d_t = l_t\, q_t + \int_t^{t+1} (s_u + r_u - w_u - e_u)\,(t+1-u)\, q_t\, d\,u$$

da cui si ricava immediatamente la (5) (1).

La determinazione è dunque rigorosa solo in quanto sia $q_{(u\,|\!-\,t+1)} = (t+1-u)\, q_t$, ciò che equivale ad ammettere che il saggio istantaneo di eliminazione diminuisca da t a $t+1$ (2). Ora tale ipotesi è per lo più in contrasto con la realtà quando si tratti di determinare la probabilità di eliminazione per morte, in quanto il saggio istantaneo di mortalità cresce col crescere dell'età dopo i primi anni di vita, del pari che con la durata dell'assicurazione, ma agli scopi pratici si assume dagli attuari come sufficientemente approssimata e in tal senso soddisfacente. Si avverte però, che, per altre cause di eliminazione, essa potrebbe indurre ad errori sensibili (3).

---

(1) Cfr. F. P. Cantelli : *Genesi e costruzione delle tavole di mutualità* in « Bollettino di Notizie sul Credito e sulla Previdenza », del Ministero dell'Agricoltura, Industria e Commercio, 1914, n. 3, pag. 37.

(2) Si avverta che facendo l'ipotesi *a*) di una distribuzione uniforme dei morti nell'intervallo $t \,|\!-\, t+1$, che sta alla base della formula di Becker, si viene ad ammettere, analogamente a quanto si ammette col metodo delle durate esatte, che la probabilità di eliminazione cresca proporzionalmente all'ampiezza della frazione dell'intervallo unitario cui si riferisce, solo che si considerano, anzichè le frazioni $u \,|\!-\, t+1$, come si fa col metodo delle durate esatte, le frazioni $t \,|\!-\, u$, ciò che porta ad ammettere che il saggio istantaneo di mortalità non diminuisca, ma cresca, come si è detto, da t a $t+1$. La formula di q_t, fondata sull'ipotesi *a*) di una distribuzione uniforme dei morti nell'intervallo $t \,|\!-\, t+1$, da applicarsi quando si conoscano le durate esatte di esposizione al rischio, è stata data dal Balducci, il quale ha anche esposto le relazioni tra detta formula e la (5) a cui si perviene nell'ipotesi *d*), col metodo delle durate esatte. La formula data dal Balducci non è però risolubile che col procedimento delle successive approssimazioni e quindi non è di comoda applicazione nella pratica. Cfr. G. Balducci: *Costruzione e critica delle tavole di mortalità*, in « Giornale degli Economisti e Rivista di Statistica », 1917, pag. 9 dell'estratto.

(3) Cfr. Cantelli : op. cit., pagg. 39-40 e 46-47.

Valga un esempio per dare un'idea della portata della ipotesi *d*) su cui il metodo delle durate esatte si fonda.

Supponiamo a tal fine che il saggio istantaneo di eliminazione resti costante, ipotesi che, secondo quanto si è detto, in molti casi risponde approssimativamente a realtà e che d'altronde è consigliabile di ammettere tutte le volte che si ignora in quale senso il saggio istantaneo di eliminazione varii nell'intervallo $t \mid\!— t + 1$. Siano 4 gruppi di esposti al rischio e si divida in quattro parti uguali l'intervallo $t \mid\!— t + 1$ della loro esistenza per il quale vogliamo determinare la probabilità di morte q_t. Sia uguale a 0.1 la probabilità di morte per ognuna delle quattro parti. In ciascuno dei quattro gruppi, vi sia una schiera di 1000 persone che entrano in osservazione al momento t e non vengono eliminate durante l'intervallo $t \mid\!— t + 1$ per altra causa che la morte, mentre un'altra schiera, costituita pure inizialmente di 1000 persone resti nel gruppo A solo in un primo quarto dell'intervallo $t \mid\!— t + 1$, per passare, alla fine del detto quarto, nel gruppo B, in cui resta durante il secondo quarto di intervallo, passando successivamente al gruppo C, dove resta durante il terzo quarto di intervallo, ed infine al gruppo D dove resta durante l'ultimo quarto di intervallo. La probabilità di morte col metodo delle durate esatte risulterebbe $= 0.3351$ per il gruppo A, $= 0.3417$ per il gruppo B, $= 0.3475$ per il gruppo C e $= 0.3528$ per il gruppo D, mentre in realtà essa è, per tutti i gruppi, $= 0.3439$, valore che si ottiene immediatamente per la schiera di esposti che non emigrano dall'uno all'altro gruppo.

Per che la probabilità determinata col metodo delle durate esatte nell'intervallo $t \mid\!— t + 1$ risultasse uguale per i 4 gruppi, converrebbe che la probabilità di morte non fosse costante, ma diminuisse nei successivi quarti di detto intervallo. Se ad esempio, essa fosse $= 0.1$ per l'ultimo quarto dell'intervallo, dovrebbe essere $= 0.111$ per il terzo, $= 0.125$ per il secondo e $= 0.143$ per il primo quarto, con una differenza quindi di oltre il 40 % tra il primo e l'ultimo quarto.

È chiaro che un'ipotesi consimile può risultare molto lontana dalla realtà e che spesso può valere meglio attenersi all'ipotesi *b*) che il saggio istantaneo di mortalità resti costante durante l'intervallo $t \mid\!— t + 1$.

Vediamo come, in base a tale ipotesi, si possa arrivare alla determinazione dei quozienti di eliminazione nel caso in cui i dati, si riferiscono tutti o in parte, a esposti in osservazione incompleta.

Il numero medio degli esposti sarà

$$M_t = l_t + \int_t^{t+1} (s_u + r_u - d_u - w_u - e_u)\,(t + 1 - u)\,d\,u$$

da cui

$$S_t = \frac{d_t}{l_t + \int_t^{t+1} (s_u + r_u - d_u - w_u - e_u)\,(t + 1 - u)\,d\,u} \qquad (6)$$

La (6) differisce dalla (5) solo per il fatto che, nel considerare la durata di esposizione al rischio, viene tenuto conto anche dell'eliminazione per la causa di cui si vuole determinare la probabilità (1).

Si avverta però che la formula (6) dà il valore del saggio medio di eliminazione S_t, mentre la (5) dà il valore della probabilità di eliminazione q_t.

Si avverta anche che, mentre la (5) è subordinata all'ipotesi *d*), la (6) non è subordinata ad alcuna ipotesi ; essa dà il valore del saggio di eliminazione valido per la massa effettiva degli esposti in osservazione sia completa che incompleta. Quando però si ammetta che il saggio istantaneo di eliminazione resta costante durante tutto l'intervallo (ipotesi *b*), il valore di S_t diventa indipendente dalla distribuzione degli esposti al rischio nei vari istanti dell'intervallo stesso e la (6) dà pertanto, in tale ipotesi, anche il valore del saggio di eliminazione che si sarebbe ottenuto nel caso in cui tutti gli esposti al rischio fossero stati in osservazione completa.

D'altra parte, nel caso in cui tutti gli esposti al rischio sieno in osservazione completa, si passa facilmente, ammettendo l'ipotesi *b*), dal valore di S_t dato dalla (6) al valore di q_t mediante la (2 *a*).

La (2 *a*) ci insegna dunque a calcolare per mezzo della (6) la probabilità di eliminazione nell'intervallo $t\,|\!-\,t+1$, nell'ipotesi *b*) che il saggio istantaneo di eliminazione si mantenga costante durante il detto intervallo.

Questo metodo potrebbe chiamarsi : *metodo delle durate esatte per saggi istantanei di eliminazione costanti*, in contrapposto al metodo usuale delle durate esatte che potrebbe dirsi : *metodo delle durate esatte per probabilità di eliminazione proporzionali.*

Dalle iniziali dei nomi degli autori che li hanno proposti designeremo questo con le lettere DEA (*metodo delle durate esatte dell'Ackland*) e quello con le lettere DEG (*metodo delle durate esatte del Gini*).

(1) La formula (6) fu già applicata nella comunicazione : *Sulla curva della fecondità matrimoniale della donna secondo l'età* presentata alla XXII riunione dell'Istituto Internazionale di Statistica (Londra, 1934).

Vediamo con qualche esempio la portata delle differenze tra i risultati ottenuti con i due metodi.

Gli esempi che seguono si riferiscono alla probabilità di concepire nei successivi anni di età tra le coniugate delle case sovrane d'Europa, delle case mediatizzate tedesche e delle case principesche d'Italia, Francia, Germania, Austria, Ungheria, Russia, Polonia e Gran Bretagna, ricordate nelle annate 1891-1927 dell'Almanacco di Gotha. Furono considerate le famiglie che hanno contratto matrimonio negli anni 1890-1909 e queste vennero osservate dalla data del loro matrimonio fino al 31 dicembre 1926. I dati vennero spogliati dall'Almanacco di Gotha dal Prof. F. Savorgnan, in occasione di alcuni articoli pubblicati nel « Metron » (Vol. III, n. 2, 1923 ; Vol. III, n. 3-4, 1924 ; Vol. IV, n. 3-4, 1925 ; Vol. V, n. 1, 1925 ; Vol. IX, n. 1, 1931), e l'elaborazione è stata eseguita nell'Istituto di Statistica della R. Università di Roma ad opera o sotto la direzione dell'Aiuto Dr. Vittorio Castellano e dell'Assistente Dr. Ferdinando Paglino.

Le coniugate sono state considerate tutte come esposte al rischio del concepimento a partire dalle nozze (non essendosi verificato nessun parto che possa attribuirsi a concepimenti prenuziali) e come eliminate da tale rischio nove mesi prima della propria morte oppure alla morte del coniuge o al momento della separazione o del divorzio. Oltre a ciò, per l'applicazione della formula (6) si sono considerate come eliminate le coniugate alla data del presumibile concepimento, vale a dire nove mesi prima del parto, mentre, per l'applicazione della formula (5), le coniugate che hanno concepito vennero considerate come eliminate alla fine dell'anno in cui il concepimento ha presumibilmente avuto luogo.

Per l'applicazione di entrambe le formule, le coniugate che hanno partorito sono state di solito considerate come nuovamente esposte al rischio del concepimento tre mesi dopo il parto, salvo in alcuni casi eccezionali in cui il parto successivo è seguito a meno di un anno di intervallo dal primo, nei quali casi esse sono state considerate come esposte al rischio nove mesi prima del secondo parto.

La tabella II indica, per i singoli anni di età, i valori di q_t e S_t in base al metodo DEG [formule (2 *a*) e (6)] e in base al metodo DEA [formula (5)]. I dati furono poi raggruppati per classi quinquennali, di età ciò che ha anche permesso di distinguere vari gruppi di case (cfr. tabelle III-VIII).

Come si vede, le differenze tra i valori di q_t ottenuti con i due metodi risultano trascurabili quando è $q_t \leqq 0.15$, ma diventano sensibili per valori superiori e tanto più notevoli quanto più q_t è elevato. Particolarmente forti risultano se si considera la probabilità del concepimento nel primo anno di matrimonio (cfr. tabella IX) : per le donne inferiori a 20 anni, la differenza supera in tal caso il 10 %. Le differenze appaiono dunque non trascurabili. Si avverta anche che esse sono sistematiche, nel senso che si verificano quasi sempre nel senso che q_t risulta inferiore col metodo DEG.

TABELLA II.

Case sovrane d'Europa, Case mediatizzate tedesche e Case principesche di Germania, Austria, Ungheria, Russia, Polonia, Italia, Francia e Gran Bretagna.

ETA'	S_t (formula 6)	q_t (formula 2 a)	q_t (formula 5)
16–18	1.1413	0.6806	0.7356
19	1.0339	0.6444	0.6957
20	0.8796	0.5851	0.6208
21	0.8551	0.5748	0.6091
22	0.5912	0.4463	0.4609
23	0.5282	0.4103	0.4146
24	0.5354	0.4146	0.4201
25	0.4310	0.3501	0.3577
26	0.4028	0.3316	0.3397
27	0.3722	0.3108	0.3108
28	0.2818	0.2456	0.2472
29	0.2821	0.2458	0.2444
30	0.2249	0.2014	0.2030
31	0.2123	0.1913	0.1926
32	0.1705	0.1568	0.1562
33	0.1129	0.1068	0.1078
34	0.1444	0.1345	0.1339
35	0.0855	0.0819	0.0841
36	0.1187	0.1119	0.1124
37	0.0793	0.0763	0.0768
38	0.0738	0.0711	0.0707
39	0.0553	0.0538	0.0539
40	0.0237	0.0234	0.0234
41	0.0389	0.0382	0.0382
42	0.0128	0.0127	0.0127
43	0.0134	0.0133	0.0133
44	0.0057	0.0057	0.0057
45	0.0061	0.0061	0.0061
46	0.0034	0.0034	0.0034
16–46	0.1894	0.1725	0.1732

TABELLA III.

Case sovrane d'Europa.

ETA'	S_t (formula 6)	q_t (formula 2 *a*)	q_t (formula 5)
16–19	1.0113	0.6363	0.6963
20–24	0.6731	0.4899	0.4996
25–29	0.4919	0.3885	0.3865
30–34	0.1955	0.1776	0.1779
35–59	0.0745	0.0718	0.0721
40–44	0.0303	0.0297	0.0297
45–46	0.0104	0.0103	0.0103
16–46	0.2206	0.1980	0.1978

TABELLA IV.

Case mediatizzate tedesche.

ETA'	S_t (formula 6)	q_t (formula 2 *a*)	q_t (formula 5)
16–19	1.0716	0.6575	0.7228
20–24	0.6513	0.4786	0.4917
25–29	0.4366	0.3538	0.3579
30–34	0.2176	0.1956	0.1968
35–39	0.1120	0.1060	0.1062
40–44	0.0172	0.0170	0.0170
45–46	0.0063	0.0063	0.0063
16–46	0.2206	0.1980	0.1983

TABELLA V.

Case principesche di Germania, Austria, Ungheria, Russia e Polonia.

ETA'	S_t (formula 6)	q_t (formula 2 a)	q_t (formula 5)
16–19	1.0137	0.6371	0.6688
20–24	0.6567	0.4814	0.4965
25–29	0.2947	0.2552	0.2585
30–34	0.1460	0.1358	0.1364
35–39	0.0770	0.0741	0.0739
40–44	0.0174	0.0172	0.0172
45–46	—	—	—
16–46	0.1789	0.1638	0.1644

TABELLA VI.

Case principesche d'Italia.

ETA'	S_t (formula 6)	q_t (formula 2 a)	q_t (formula 5)
16–19	1.0461	0.6487	0.7325
20–24	0.5741	0.4368	0.4526
25–29	0.2553	0.2253	0.2255
30–34	0.1017	0.0967	0.0961
35–39	0.0536	0.0522	0.0523
40–44	0.0182	0.0180	0.0181
45–46	—	—	—
16–46	0.1407	0.1312	0.1305

TABELLA VII.

Case principesche di Francia.

ETA'	S_t (formula 6)	q_t (formula 2 *a*)	q_t (formula 5)
16–19	1.3990	0.7532	0.8176
20–24	0.5632	0.4306	0.4390
25–29	0.2429	0.2156	0.2143
30–34	0.1504	0.1396	0.1393
35–39	0.0663	0.0641	0.0646
40–44	0.0137	0.0136	0.0136
45–46	—	—	—
16–46	0.1781	0.1631	0.1632

TABELLA VIII.

Case sovrane d'Europa, Case mediatizzate tedesche e Case principesche di Germania, Austria, Ungheria, Russia, Polonia, Italia, Francia e Gran Bretagna.

ETA'	S_t (formula 6)	q_t (formula 2 *a*)	q_t (formula 5)
16–19	1.0711	0.6574	0.7099
20–24	0.6275	0.4661	0.4791
25–29	0.3413	0.2892	0.2915
30–34	0.1715	0.1576	0.1580
35–39	0.0829	0.0796	0.0797
40–44	0.0195	0.0193	0.0193
45–46	0.0048	0.0048	0.0048
16–46	0.1894	0.1725	0.1732

TABELLA IX.

Case sovrane d'Europa, Case mediatizzate tedesche e Case principesche di Germania, Austria, Ungheria, Russia, Polonia, Italia, Francia e Gran Bretagna.

Concepimenti avvenuti durante il 1° anno di matrimonio.

ETA'	S_t (formula 6)	q_t (formula 2 *a*)	q_t (formula 5)
15–19	1.5724	0.7925	0.8949
20–24	1.4436	0.7639	0.8378
25–29	0.9909	0.6288	0.6769
30–34	0.5507	0.4235	0.4356
35–39	0.3962	0.3271	0.3068

* * *

Nella tabella II noi abbiamo raggruppato le età da 16 a 18 anni. Se le avessimo invece distinte, avremmo ottenuto i risultati seguenti (cfr. tabella X), i quali mettono in luce un inconveniente del metodo DEA delle durate esatte, in quanto esso può condurre a valori di q_t superiori all'unità.

TABELLA X.

ETÀ	S_t (formula 6)	q_t (formula 2 *a*)	q_t (formula 5)
16.	10.4712	0.9999	5.2910
17.	1.8767	0.8469	0.8264
18.	0.9674	0.6200	0.6666

L'inconveniente non è solo teorico, ma anche pratico, e può verificarsi pure in osservazioni molto numerose quando molti esposti al rischio entrino in osservazione nell'ultima parte dell'in-

tervallo $t \,|\!-\, t+1$ e quando sia contemporaneamente elevata la probabilità di eliminazione. Supponiamo, ad esempio, che tutti gli esposti al rischio entrino in osservazione solo dopo che sono trascorsi 9/10 dell'intervallo $t \,|\!-\, t+1$ e che la probabilità di eliminazione durante l'ultimo decimo dell'intervallo sia $q_{(t+0.9\,|\!-\,t+1)} = 0.42$. Col metodo DEA. delle durate esatte si ottiene $q_t = 10\, q_{(t+0.9\,|\!-\,t+1)} = 4.2$. Un caso simile si verifica appunto per la probabilità di eliminazione per concepimento delle coniugate per l'anno più giovanile (nel nostro esempio, per l'anno 16°), in quanto la probabilità del concepimento è molto elevata e generalmente le coniugate si sposano quando sono vicine al limite superiore dell'anno. È pertanto da ritenere che il risultato esposto nella tabella X non sia dovuto al numero ristretto delle osservazioni.

Il valore del saggio di eliminazione S_t può naturalmente risultare superiore all'unità ; ma, comunque sia alto S_t, il valore di q_t ottenuto dalla (2 *a*) risulta sempre < 1. Nel nostro esempio, per l'anno 16°, risulta $q_t = 0.9999$. Il metodo DEG delle durate esatte non presenta dunque l'accennato inconveniente del metodo DEA.

* * *

Sarà bene mettere in luce la portata delle differenze tra i risultati ottenuti col metodo qui proposto e quelli che si otterrebbero col metodo usuale per misurare la fecondità delle donne maritate.

Si usano, a tal fine, come è noto, i coefficienti specifici di fecondità matrimoniale, che si ottengono ragguagliando il numero dei parti verificatisi durante un anno tra le donne di una data età (o il numero delle nascite qualora si voglia contare un parto tante volte quanti sono i nati a cui ha dato luogo) al numero medio delle donne coniugate in tale età, esistenti durante l'anno. Si considerano, pertanto, come esposte al rischio le donne maritate dal momento delle nozze al momento dello scioglimento del matrimonio avvenuto per morte della moglie o del marito, o per divorzio, o per separazione.

Tale metodo differisce dal nostro per tre circostanze :

α) perchè le donne sono considerate come esposte al rischio del parto fin dal momento delle nozze, mentre (a parte il caso di concezioni prenuziali, che, nell'esempio in esame, non si verifica), esse da tale momento sono in realtà esposte al rischio del concepi-

mento, ed al rischio del parto sono normalmente esposte solo dopo nove mesi di matrimonio;

β) perchè, anche in seguito, sono considerate come esposte al rischio del parto tutte le donne coniugate, senza cioè fare eccezione per quelle che hanno da poco partorito e che non potranno partorire ulteriormente se non dopo parecchi mesi (noi abbiamo generalmente ammesso un termine medio di dodici mesi supponendo che le persone saranno nuovamente esposte al rischio del concepimento tre mesi dopo il parto, salvo quando risultava un periodo più breve) ;

γ) perchè sono considerate come eliminate dal rischio del parto le coniugate alla morte del marito, o al divorzio, o alla separazione, mentre a tale momento cessano di essere esposte al rischio del concepimento, ma restano esposte per altri nove mesi al rischio del parto.

Nella colonna *A* della tabella seguente (XI) sono esposti i coefficienti specifici di fecondità, calcolati secondo il metodo usuale per il complesso delle famiglie sovrane, delle case mediatizzate tedesche e delle case principesche dei vari Stati. Tali coefficienti possono mettersi a confronto con i saggi di eliminazione per concepimento, calcolati secondo la (6) e riprodotti nella colonna *D*.

Al fine di renderci conto separatamente dell'influenza delle circostanze α), β) e γ) sono stati calcolati :

1) i saggi di eliminazione per concepimento, considerando come esposte al rischio tutte le donne maritate dal momento delle nozze fino allo scioglimento del matrimonio. Tali saggi si trovano esposti nella colonna *C*. Le differenze tra i dati delle colonne *D* e *C* mostrano dunque la portata della circostanza β), la quale appare molto notevole

2) I saggi di eliminazione per parto, considerando come eliminate dal rischio del parto le coniugate alla loro morte, oppure nove mesi dopo la morte del marito, la separazione o il divorzio. Tali saggi si trovano esposti nella colonna *B*. Le differenze tra i dati delle colonne *A* e *B* mostrano dunque la portata della circostanza γ) la quale appare trascurabile.

Infine le differenze tra i dati delle colonne *B* e *C* mostrano la portata della circostanza α) la quale appare molto notevole.

Nella successiva tabella XII i risultati dei calcoli sono condensati per classi quinquennali di età : spariscono in essa le oscillazioni dovute al ristretto numero delle osservazioni e meglio risal-

TABELLA XI.

Case sovrane, Case mediatizzate tedesche e Case principesche di vari Stati.

ETÀ	Valori di S_t calcolati col procedimento			
	A	*B*	*C*	*D*
16–18	0.2615	0.2615	0.6755	1.1413
19	0.3177	0.3157	0.6031	1.0339
20	0.4088	0.4031	0.5064	0.8796
21	0.3712	0.3704	0.4980	0.8551
22	0.3950	0.3950	0.3640	0.5912
23	0.3258	0.3241	0.3443	0.5282
24	0.3275	0.3260	0.3540	0.5354
25	0.2842	0.2817	0.2978	0.4310
26	0.2835	0.2820	0.2840	0.4028
27	0.2870	0.2859	0.2415	0.3722
28	0.2312	0.2293	0.2164	0.2818
29	0.2370	0.2365	0.2150	0.2821
30	0.1813	0.1806	0.1798	0.2249
31	0.1762	0.1745	0.1715	0.2123
32	0.1849	0.1835	0.1393	0.1705
33	0.1153	0.1140	0.0983	0.1129
34	0.1121	0.1112	0.1233	0.1444
35	0.1115	0.1102	0.0755	0.0855
36	0.0731	0.0720	0.1031	0.1187
37	0.1088	0.0071	0.0721	0.0793
38	0.0894	0.0886	0.0657	0.0738
39	0.0449	0.0445	0.0508	0.0553
40	0.0559	0.0552	0.0220	0.0237
41	0.0230	0.0228	0.0366	0.0389
42	0.0290	0.0287	0.0120	0.0128
43	0.0179	0.0177	0.0128	0.0134
44	0.0110	0.0109	0.0054	0.0057
45	0.0058	0.0057	0.0058	0.0061
46	0.0032	0.0031	0.0032	0.0034
47	0.0035	0.0035	—	—
16–47 (1)	0.1538	0.1525	0.1557	0.1894

(1) 16-46 per le colonne *C* e *D*.

TABELLA XII.

Case sovrane, Case mediatizzate tedesche e Case principesche di vari Stati.

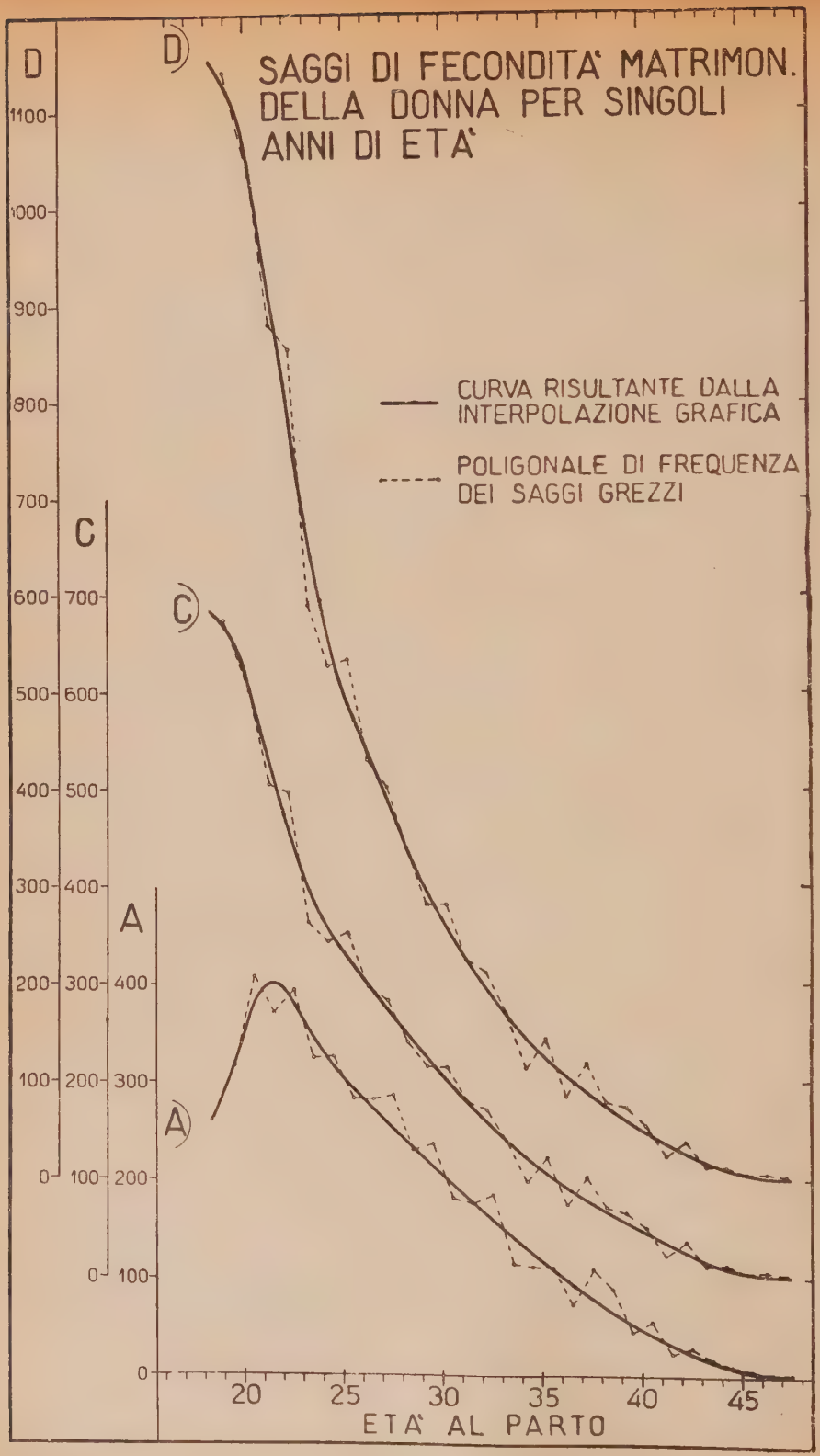
ETÀ	Valori di S_t calcolati col procedimento			
	A	*B*	*C*	*D*
16–19.	0.2985	0.2972	0.6279	0.0711
20–24.	0.3571	0.3556	0.3936	0.6275
25–29.	0.2634	0.2619	0.2489	0.3413
30–34.	0.1539	0.1527	0.1423	0.1715
35–39.	0.0861	0.0851	0.0739	0.0829
40–44.	0.0283	0.0280	0.0184	0.0195
45–46 (1)	0.0042	0.0042	0.0046	0.0048
16–46 (2)	0.1538	0.1525	0.1557	0.1894

(1) 45-47 per le colonne *A* e *B*.
(2) 16-47 per le colonne *A* e *B*.

tano quindi le divergenze sistematiche fra i risultati ottenuti con vari metodi.

Conviene però tener presente, nei confronti, che, mentre i saggi indicati nelle colonne *A* e *B* si riferiscono all'età della madre al parto, i saggi indicati nelle colonne *C* e *D*, si riferiscono alla sua età al concepimento, che, normalmente, precede di nove mesi la prima. È essenzialmente per questo che, al di sopra di una certa età, i dati della colonna *B* risultano più elevati dei dati della colonna *C* per le stesse età.

Ad ovviare a tale difficoltà e a permettere di rendersi conto a colpo d'occhio dell'influenza delle circostanze α) e β) furono costruiti i diagrammi qui uniti, corrispondenti ai dati delle colonne *A*, *C* e *D* trasportando di nove mesi verso destra le due ultime curve. Sull'asse delle ascisse risulta quindi indicata l'età al parto. L'età media della prima classe (al di sotto di 19 anni) corrisponde ad anni 18.2 per i concepimenti (curve *C* e *D*) e ad anni 18.3 per i parti (curva *A*). Parve inutile di costruire anche la curva relativa ai dati della colonna *B*, dato che questi poco differiscono dai dati della colonna *A*. Per eliminare le oscillazioni dipendenti dal numero ristretto delle osservazioni, fu eseguita altresì un'interpola-


SAGGI DI FECONDITA' MATRIMON. DELLA DONNA PER SINGOLI ANNI DI ETA'
CURVA RISULTANTE DALLA INTERPOLAZIONE GRAFICA
POLIGONALE DI FREQUENZA DEI SAGGI GREZZI
D
C
A
1100
1000
900
800
700
600
500
400
300
200
100
0
20
25
30
35
40
45
ETA' AL PARTO

zione grafica, avendo cura che, per ciascuna curva, i saggi interpolati corrispondessero a un numero complessivo di nati uguale a quello effettivo (1).

I diagrammi mettono in luce come il non tener conto della circostanza β) abbia per effetto di ridurre le differenze tra i saggi di eliminazione relativi alle successive classi di età, attenuando così l'inclinazione della curva, senza però modificarne la forma [vedi curva *C* in confronto alla curva *D*]. Il non tener conto della circostanza α) ha invece per effetto di abbassare i saggi di eliminazione delle nuove spose, più frequenti nelle età più giovani, facendo risultare i saggi, fino a 20 anni, più bassi anzi che più alti di quelli per le età successive. Ne risulta, pertanto, modificata la forma della curva che presenta il massimo nella classe di età di 20-24 anni (e precisamente secondo i dati interpolati, verso 21 anni), anzi che nella classe di età inferiori a 20 anni [vedi curva *A* in confronto alla *C*] (2).

Nella tabella XIII infine sono riportati i dati interpolati in corrispondenza all'età al parto secondo le tre curve (col. 1, 2 e 3) e in corrispondenza all'età al concepimento secondo le curve *C* e *D* (col. 4 e 5).

* * *

Quando il numero delle osservazioni è molto grande, l'applicazione del metodo delle durate esatte viene riguardato come troppo faticoso, cosicchè gli attuari preferiscono ricorrere a metodi abbreviati pure proposti dall'Ackland e noti sotto i nomi di *metodo delle durate medie* e *metodo delle durate più vicine*.

Entrambi suppongono che la distribuzione degli entrati e degli usciti, o quanto meno la distribuzione delle differenze tra questi e quelli, sia uniforme durante tutto l'intervallo $t \mid\!\!- t + 1$.

---

(1) La interpolazione è stata eseguita dalla Dott.ssa E. Rossi, studente della Scuola di Scienze Statistiche e Attuariali della R. Università di Roma.

(2) Questi effetti delle circostanze α) e β) erano già stati esattamente previsti nella nostra comunicazione: *Sur la mesure de la fécondité des mariages* (« Bulletin de l'Institut International de Statistique », XXI Sessione, Messico 1933, Tomo XXVII, pp. 37 e 38 dell'estratto), in cui venivano segnalati gli inconvenienti che derivano dagli usuali coefficienti specifici di fecondità matrimoniale. Nella successiva comunicazione: *Sulla curva della fecondità matrimoniale della donna secondo l'età*, già citata, i valori dei coefficienti specifici di fecondità matrimoniale (procedimento *A*) erano posti a confronto coi valori dei sagg di eliminazione ottenuti con la formula (6) (procedimento *D*, in detta comunicazione indicato con la lettera *c*).

TABELLA XIII.

Case sovrane, Case mediatizzate tedesche e Case principesche di vari Stati.

ETÀ AL PARTO	Valori interpolati di S_t calcolati col procedimento			ETÀ AL CONCEPIMENTO	Valori interpolati di S_t calcolati col procedimento	
	A	*C*	*D*		*C*	*D*
	1	2	3		4	5
16–18	0.2615	0.6876	1.1564	16–18	0.6755	1.1413
19	0.3240	0.6540	1.1126	19	0.6102	1.0472
20	0.3876	0.5898	1.0140	20	0.5300	0.9106
21	0.4042	0.5118	0.8842	21	0.4628	0.7930
22	0.3845	0.4460	0.7524	22	0.3992	0.6612
23	0.3478	0.3878	0.6318	23	0.3588	0.5626
24	0.3196	0.3514	0.5372	24	0.3290	0.4935
25	0.2960	0.3220	0.4786	25	0.3022	0.4428
26	0.2744	0.2962	0.4296	26	0.2760	0.3936
27	0.2556	0.2700	0.3840	27	0.2506	0.3434
28	0.2360	0.2442	0.3308	28	0.2256	0.2966
29	0.2178	0.2200	0.2876	29	0.2046	0.2622
30	0.1992	0.1994	0.2514	30	0.1824	0.2304
31	0.1812	0.1760	0.2208	31	0.1610	0.1984
32	0.1638	0.1558	0.1906	32	0.1414	0.1705
33	0.1438	0.1360	0.1638	33	0.1236	0.1462
34	0.1276	0.1198	0.1416	34	0.1084	0.1260
35	0.1115	0.1054	0.1214	35	0.0942	0.1078
36	0.0960	0.0916	0.1040	36	0.0804	0.0918
37	0.0814	0.0780	0.0878	37	0.0682	0.0760
38	0.0676	0.0662	0.0724	38	0.0576	0.0620
39	0.0558	0.0548	0.0594	39	0.0464	0.0500
40	0.0442	0.0440	0.0470	40	0.0368	0.0378
41	0.0344	0.0342	0.0358	41	0.0266	0.0280
42	0.0260	0.0246	0.0256	42	0.0180	0.0184
43	0.0182	0.0101	0.0166	43	0.0104	0.0110
44	0.0112	0.0096	0.0102	44	0.0064	0.0070
45	0.0058	0.0058	0.0062	45	0.0042	0.0044
46	0.0032	0.0038	0.0040	46	0.0026	0.0027
47	0.0022	0.0020	0.0022			

In tal caso nella (5) si può porre $t + 1 - u = \frac{1}{2}$, e si ottiene allora la formula:

$$ {}_m q_t = \frac{d_t}{l_t + \frac{K_t}{2}} \qquad (5\ m) $$

dove è

$$ K = \Sigma\ (s_u + r_u - w_u - e_u)\ . $$

La (5 *m*) dà appunto la probabilità di eliminazione col metodo delle durate medie.

Col metodo delle durate più vicine, invece, si pone $u = t$ o $u = t + 1$ a seconda che u sia più vicino a t o a $t + 1$.

Applicando ai nostri dati il metodo delle durate medie (formula 5 *m*), si ottengono i valori indicati nella colonna 2 della tabella XVI che sono posti a confronto coi valori della colonna 1, già ottenuti col metodo delle durate esatte (formula 5). Tra le due serie di valori non si avvertono divergenze sistematiche: 11 volte infatti i valori ottenuti col metodo delle durate medie risultano superiori, 2 volte uguali fino al 4° decimale, 16 volte inferiori a quelli ottenuti col metodo delle durate esatte. I valori determinati col metodo delle durate medie presentano però maggiori irregolarità, spiegabili anche col numero ristretto delle osservazioni; per le età più giovani, le differenze cadono sulla seconda cifra decimale e possono pertanto non risultare trascurabili ai fini di talune indagini.

Una semplificazione analoga può apportarsi alla determinazione di S_t mediante la (6), facendo $t + 1 - u = \frac{1}{2}$ e ottenendo la formula:

$$ {}_m S_t = \frac{d_t}{l_t + \frac{K_t + d_t}{2}} \qquad (6\ m) $$

dalla quale poi, in base alla (2 *a*), si può risalire al valore di ${}_m q_t$ della probabilità di eliminazione.

Le colonne 4 e 7 della tabella XVI danno i valori di ${}_m S_t$ e di ${}_m q_t$, così ottenuti, che si possono confrontare coi rispettivi valori di S_t secondo la (6) e di q_t secondo il metodo DEG, riprodotti nelle colonne 3 e 6.

Le differenze tra i valori riportati nelle colonne 3 e 4 e quelle, che ne seguono, fra i valori delle colonne 6 e 7, si verificano, salvo in due casi, nello stesso senso che fra i valori delle colonne 1 e 2 e sono dello stesso ordine di grandezza. I valori ottenuti mediante la formula (6 *m*) (riportati nelle colonne 4 e 7) 13 volte risultano superiori, 2 volte uguali e 14 inferiori a quelli ottenuti con la formula (6) (riportati nelle colonne 3 e 6).

Ne segue che le differenze tra i valori della col. 2 ottenuti col metodo delle durate medie dell'Ackland e quelli della colonna 7 ottenuti con la (2 *a*) attraverso la (6 *m*) ,risultano analoghe per senso e per intensità alle differenze tra i valori della colonna 1, ottenuti col metodo (DEA) delle durate esatte dell'Ackland e quelli della colonna 6, ottenuti col nostro metodo (DEG) delle durate esatte. Anche tra le due serie di valori delle col. 2 e 7 le differenze risultano trascurabili quando è $q_t \lesssim 0{,}15$.

* * *

I metodi delle durate medie e delle durate più vicine non costituiscono che la semplificazione del metodo delle durate esatte e, come questo, sono soggetti all'ipotesi *d*). Poichè è riconosciuto che tale ipotesi, in molti casi, non risponde al vero, si è studiato un procedimento che conduca a determinare i valori di q_t su ipotesi che permettano di avvicinarsi maggiormente alla realtà, supponendo, cioè, che il numero degli eliminati varii in funzione lineare del tempo durante l'intervallo $t \mid\!— t + 1$ [ipotesi *e*)] (1). Le semplificazioni fatte di detto procedimento hanno confermato la prevista insufficienza, in casi particolari, del metodo delle durate medie ed hanno permesso di mettere in luce la portata delle inesattezze a cui questo può portare, ma il procedimento è talmente laborioso da non essere consigliabile in pratica, ed effettivamente, all'infuori delle esemplificazioni su ricordate, esso non fu mai applicato. Noi ne diamo un'altra esemplificazione sui nostri dati allo scopo di mostrare le divergenze tra i risultati che con esso si ottengono e quelli ottenuti con le formule precedentemente esposte.

Nel procedimento in parola, si comincia col determinare i valori di ${}_m q_t$ col metodo delle durate medie dell'Ackland (formula 5 *m*). Tali valori, nel nostro caso, sono riportati nella colonna 5

---

(1) Cfr. BALDUCCI: art. cit.

della tabella XIV. Ai valori così ottenuti si aggiunge un termine di correzione $-\frac{{}_m q_t^3}{6}\frac{K_t}{d_t}$ (cfr. col. 6, tab. XIV), al fine di determinare approssimativamente i valori di ${}_m q_t$ (cfr. col. 7, tab. XIV) che si sarebbero ottenuti nell'ipotesi che gli eliminati, anzi che distribuirsi secondo l'ipotesi *d*), si distribuissero, secondo l'ipotesi *a*), uniformemente in ciascun intervallo $t \mid\!- t+1$ (nel nostro caso in ciascun anno di età) (1). Su tali valori ci si basa per determinare i valori di ${}_m q_t$ che si otterrebbero nell'ipotesi che gli eliminati, anzichè distribuirsi uniformemente in ciascun intervallo, variassero durante esso in funzione lineare del tempo (ipotesi *e*). Naturalmente il valore del coefficiente angolare della retta rappresentativa di tale funzione varia di intensità nei successivi intervalli considerati (nel caso nostro, nei successivi anni di età). Al fine di determinare tale valore per i successivi intervalli, conviene, quando il numero delle osservazioni non sia molto elevato (e questo appunto è il caso nostro), procedere preliminarmente ad una perequazione dei valori di ${}_m q_t$ ottenuti nell'ipotesi *a*). La colonna 8 contiene i valori ottenuti con una perequazione grafica, mentre nella successiva colonna 9 sono esposti i valori β_t dei coefficienti angolari rispettivi (2). Essi ci permettono di determinare i valori della probabilità di eliminazione nell'ipotesi *e*), in base alla formula :

$$ {}_m q_t = \frac{d_t}{l_t + \left(1 + \frac{\beta_t}{6}\right)\frac{K_t}{2}} \qquad (7\ m) $$

Tali valori, esposti nella colonna 10, sono però affetti da errori, i quali si possono determinare approssimativamente mediante la formula $-\frac{{}_m q_t^3}{6}\frac{K_t}{d_t}\left(1 - \frac{\beta_t^2}{20}\right)$. Tenendo conto dei termini di correzione forniti da tale formula (cfr. col. 11), si ottengono infine i valori definitivi della probabilità di eliminazione (cfr.

---

(1) È l'ipotesi che sta a base della formula del Wittstein, della quale i valori della col. 7 rappresentano approssimazioni sufficienti agli scopi pratici.

(2) Il valore β_t per l'intervallo $t \mid\!- t+1$ è stato ricavato con la formula :

$$ \beta_t = \frac{{}_m q_{t+1}(1 - {}_m q_t)(1 - {}_m q_{t-1}) - {}_m q_{t-1}}{2\,{}_m q_t (1 - {}_m q_{t-1})} $$

Cfr. Balducci : art. cit., formula (60).

ETÀ	Valori di l_t	Valori di d_t	Valori di Kt	Valori di ${}_m q_t$ calcolati col metodo delle durate medie (formula 5 m)	Termini di correzione	Valori di ${}_m q_t$ nel ipotesi che gli el minati si distribu scano uniforme- mente nel tempo
1	2	3	4	5	6	7
18	10	31	62	0.75610	— 0.14406	0.61204
19	31	53	87	0.71140	— 0.09849	0.61291
20	65	77	125	0.60392	— 0.05959	0.54433
21	113	116	153	0.61213	— 0.05141	0.56072
22	150	108	175	0.45473	— 0.02541	0.42932
23	217	119	141	0.41391	— 0.014	0.39991
24	239	139	170	0.42901	— 0.01609	0.41292
25	270	129	175	0.36083	— 0.01061	0.35022
26	318	130	148	0.33333	— 0.00702	0.32631
27	334	117	151	0.28571	— 0.00501	0.28070
28	368	108	133	0.24856	— 0.00314	0.24542
29	393	111	123	0.24422	— 0.00268	0.24154
30	405	95	122	0.20386	— 0.00181	0.20205
31	432	92	94	0.19210	— 0.00121	0.19089
32	434	75	94	0.15590	— 0.00079	0.15511
33	453	53	71	0.10849	— 0.00028	0.10821
34	471	66	53	0.13266	— 0.00031	0.13235
35	458	40	50	0.08281	— 0.00011	0.08270
36	468	53	24	0.11041	— 0.00010	0.11031
37	439	36	43	0.07809	— 0.00009	0.07800
38	446	32	21	0.06994	— 0.00003	0.06991
39	436	24	14	0.05417	— 0.000014	0.05416
40	426	10	8	0.02325	— 0.000001	0.02325
41	424	16	— 15	0.03841	+ 0.000007	0.03841
42	393	5	4	0.01265	— 0.0000002	0.01265
43	392	5	— 29	0.01324	+ 0.0000017	0.01324
44	358	2	— 11	0.00567	+ 0.00000005	0.00567
45	345	2	— 30	0.00606	+ 0.0000004	0.00606
46	313	1	— 27	0.00333	+ 0.000001	0.00333

(+) Tra le spose di 18 anni sono state comprese anche le poche in età di 16 e 17.

ase principesche di vari Stati. TABELLA XIV (*).

Valori a col. 7 requati	Valori di β_t dedotti dai valori della col. 8	Valori di $_m q_t$ calcolati in prima approssimazione nella ipotesi che gli eliminati si distribuiscano secondo una funzione lineare del tempo	Termini di correzione	Valori di $_m q_t$ calcolati in seconda approssimazione nella ipotesi della col. 10	Valori di $_m q_t$ calcolati nella ipotesi di saggi istantanei di mortalità costanti (formula $2a$ attraverso la $6m$)
8	9	10	11	12	13
.6213	—	—	—	—	0.70350
.6110	— 1.15822	0.80178	— 0.13155	0.67023	0.66852
.5792	— 1.16295	0.66732	— 0.07497	0.59235	0.57902
.5312	— 1.08767	0.66047	— 0.05959	0.60088	0.58610
.4711	— 0.96405	0.48335	— 0.02914	0.45421	0.44487
.4250	— 0.78619	0.42765	— 0.01496	0.41269	0.40662
.3869	— 0.67631	0.44209	— 0.01721	0.42488	0.42083
.3521	— 0.60182	0.36992	— 0.01124	0.35868	0.35614
.3199	— 0.54137	0.33914	— 0.00729	0.33185	0.32968
.2898	— 0.48085	0.29000	— 0.00518	0.28482	0.28345
.2699	— 0.44376	0.25141	— 0.00323	0.24818	0.24708
.2308	— 0.46241	0.24680	— 0.00274	0.24406	0.24285
.2031	— 0.39105	0.20562	— 0.00185	0.20377	0.20308
.1772	— 0.35996	0.19321	— 0.00122	0.19199	0.19144
.1547	— 0.33106	0.15677	— 0.00079	0.15598	0.15558
.1336	— 0.30847	0.10890	— 0.00028	0.10862	0.10837
.1161	— 0.27885	0.13299	— 0.00031	0.13268	0.13247
.1012	— 0.25556	0.08299	— 0.00011	0.08288	0.08277
.0886	— 0.24300	0.11053	— 0.00010	0.11043	0.11033
.0763	— 0.24302	0.07832	— 0.00010	0.07822	0.07808
.0651	— 0.24597	0.07001	— 0.00004	0.06997	0.06993
.0541	— 0.26848	0.05421	— 0.00002	0.05419	0.05418
.0429	— 0.31074	0.02327	— 0.00002	0.02325	0.02323
.0319	— 0.37324	0.03837	— 0.000008	0.03837	0.03844
.0217	— 0.48562	0.01266	— 0.0000002	0.01266	0.01262
.0118	— 0.63821	0.01319	+ 0.000002	0.01319	0.01321
.0072	— 0.49157	0.00567	+ 0.0000002	0.00567	0.00568
.0049	— 0.32407	0.00605	+ 0.0000005	0.00605	0.00608
.0041	—	—	—	—	0.00330

(*) I calcoli necessari per la compilazione di questa tabella sono stati eseguiti lla Dott.ssa E. Rossi, studente della Scuola di Scienze statistiche e attuariali della . Università di Roma.

col. 12). La col. 12 ci dà dunque i valori della probabilità di eliminazione nell'ipotesi che i numeri degli entrati e degli usciti (o quanto meno le loro differenze) si distribuiscano uniformemente nell'intervallo e che gli eliminati si distribuiscano invece secondo una funzione lineare del tempo ; nel caso nostro, che essi diminuiscano col crescere dell'età.

Il confronto tra i dati delle colonne 5 e 7 mostra come il metodo delle durate medie in base all'ipotesi *d*) porti, in confronto al metodo delle durate medie in base all'ipotesi *a*), a valori superiori od inferiori di ${}_mq_t$, a seconda che è $K_t \gtrless 0$; ed il confronto dei dati delle dette col. 5 e 7 coi dati della colonna 13 mostra come il metodo delle durate medie in base all'ipotesi *b*) porti sempre a valori intermedi tra i due precedenti. Per rendersi conto di tali risultati, si consideri che, quando K_t è positivo e la distribuzione degli entrati e degli usciti (o delle loro differenze) è uniforme, come si suppone nel metodo delle durate medie, gli individui in osservazione (a prescindere dagli eliminati) crescono dal principio alla fine dell'intervallo $t \mid\!- t+1$, mentre il contrario avviene quando K_t è negativo. Ora, secondo l'ipotesi *a*), il saggio di mortalità va crescendo ; secondo l'ipotesi *d*), invece, esso va diminuendo nel corso dell'intervallo $t \mid\!- t+1$, per modo che i K_t individui, i quali entrano od escono durante l'intervallo, corrispondono ad un numero di individui osservati durante tutto l'intervallo, che risulta maggiore secondo l'ipotesi *a*) che secondo l'ipotesi *d*). Ne discende che, a seconda che K_t è positivo o negativo, il denominatore della frazione che dà il valore di ${}_mq_t$ risulta maggiore o rispettivamente minore secondo l'ipotesi *a*) che secondo l'ipotesi *d*) ; e quindi il valore di ${}_mq_t$ risulta minore o rispettivamente maggiore secondo l'ipotesi *a*) che secondo l'ipotesi *d*). L'ipotesi *b*), che presuppone un andamento dei saggi di mortalità intermedio tra quelli che corrispondono alle ipotesi *a*) e *d*), porta necessariamente a valori intermedi di ${}_mq_t$.

I valori di ${}_mq_t$, ottenuti in prima approssimazione in base all'ipotesi *e*) (col. 10), portano a valori di ${}_mq_t$ superiori o inferiori a quelli ottenuti nell'ipotesi *d*) (col. 5), a seconda che β_t è positiva o negativa, come si intende subito, confrontando la formula (5 *m*) con la (7 *m*) ; ma il successivo termine di correzione (salvo per valori di ${}_mq_t$ molto bassi) risulta negativo (col. 11), per modo che i valori di ${}_mq_t$, ottenuti in seconda approssimazione nell'ipotesi *e*) (col. 12), risultano spesso intermedi tra i valori ottenuti in base

alle ipotesi *a*) e *d*). Precisamente essi risultano intermedi per i valori di ${}_mq_t$ fino al 31° anno di età compreso, mentre per gli anni successivi, risultano superiori per K_t positivo e inferiori per K_t negativo, ma per differenze piccolissime.

I valori ottenuti con le lunghe elaborazioni rese necessarie quando si adotta la ipotesi *e*), non differiscono sostanzialmente da quelli ottenuti, molto più rapidamente, in base all'ipotesi *b*. La differenza massima tra le due serie di valori risulta invero di 0.014, mentre essa risulta di 0.041 tra i valori ottenuti nell'ipotesi *a*) e quelli ottenuti nell'ipotesi *e*) e di 0.057 tra i valori ottenuti nell'ipotesi *d*) e quelli ottenuti nell'ipotesi *e*).

Questa esemplificazione conferma la praticità della determinazione delle probabilità di eliminazione in base all'ipotesi *b*).

Con ciò non si intende, naturalmente, che l'ipotesi *b*) ed i metodi che su di essa si basano siano applicabili in ogni caso. Vi sono casi, come nella mortalità durante i primi anni di vita, in cui si conosce che i saggi di eliminazione diminuiscono rapidamente e in cui (esclusa la convenienza di ricorrere alle laboriose elaborazioni in base all'ipotesi *e*)), sarà quindi opportuno ricorrere, anzichè all'ipotesi *b*), all'ipotesi *d*) che sta a base dei metodi dell'Ackland. In altri casi, in cui si conosca che i saggi di eliminazione vanno crescendo, potrà essere consigliabile ricorrere all'ipotesi *a*). Molte altre volte, però, è certo che l'ipotesi *d*) dei saggi di eliminazione costanti si avvicina sufficientemente al vero, così da giustificare l'applicazione dei metodi proposti in questa memoria, che su di essa si fondano.

* * *

Si potrebbe osservare che la semplificazione $t + 1 - u = \frac{1}{2}$ che abbiamo introdotto nella (6) è, a rigore, contraddittoria coll'ipotesi della costanza dei saggi istantanei di eliminazione che sta alla base del metodo DEG, in quanto essa suppone che nell'intervallo $t \,|\!-\, t + 1$ si distribuiscano uniformemente, non solo gli entrati e gli usciti, ma anche gli eliminati, mentre, se i saggi istantanei di eliminazione sono costanti, il numero degli eliminati tra gli esposti al rischio all'inizio t dell'intervallo diminuisce col tempo proporzionalmente al diminuire di tali esposti al rischio, mentre conviene, d'altra parte, aggiungere gli eliminati provenienti dagli entrati e togliere quelli che sarebbero provenuti dagli usciti.

In realtà l'obbiezione sarebbe infondata, in quanto la semplificazione $t + 1 - u = \frac{1}{2}$ non suppone precisamente che nell'intervallo $t \,|\!-\, t + 1$ si distribuiscano uniformemente gli entrati, gli usciti e gli eliminati, ma solo che si distribuiscano uniformemente le differenze tra i numeri degli entrati e degli usciti o eliminati, ipotesi che non è contraddittoria con l'altra che il numero degli eliminati varii proporzionalmente al numero degli esposti al rischio.

D'altra parte anche l'ipotesi che gli entrati e gli usciti (o quanto meno le loro differenze) siano distribuiti uniformemente nell'intervallo, molte volte non risponde a verità: in taluni casi pare più plausibile ammettere che il numero degli entrati e degli usciti (o quanto meno la loro differenza) sia proporzionale al numero degli esposti al rischio secondo la legge di eliminazione, e a tale ipotesi si ricorre effettivamente, come è noto, da qualche autore (Zeuner) nella trattazione delle probabilità di eliminazione.

Vediamo dunque quale espressione assuma la (5) quando si supponga, da una parte che i saggi istantanei di eliminazione sieno costanti e dall'altra che la differenza tra entrati ed usciti varii durante l'intervallo $t \,|\!-\, t + 1$ proporzionalmente agli esposti al rischio secondo la legge di eliminazione.

Se con Δ si indica la durata media di esposizione al rischio degli entrati, Δ indicherà anche la durata media dell'intervallo durante il quale restano esclusi dal rischio sia gli usciti sia i d'_t eliminati provenienti dagli l_t individui entrati al principio dell'intervallo. Si trova facilmente (1)

$$\Delta = \frac{1}{1 - e^{-S_t}} - \frac{1}{S_t} = \frac{1}{q_t} - \frac{1}{S_t}$$

---

(1) Se analogamente a quanto si è fatto a pagg. 3-4, si divide l'intervallo unitario $t \,|\!-\, t + 1$ in un numero n di intervalli istantanei uguali tra di loro, così piccoli da poter considerare senza errore sensibile gli entrati in (o gli usciti di) osservazione in ciascuno di essi come entrati (o usciti) all'inizio del rispettivo intervallo, la durata complessiva di esposizione al rischio degli entrati nell'intervallo i^{esimo} (e la durata di *non* esposizione al rischio per gli usciti), nella ipotesi del testo (che il numero degli entrati (e degli usciti) sia proporzionale al numero degli esposti al rischio), risulta proporzionale a $\frac{n-1}{n}\left(1 - \frac{S_t}{n}\right)^i$, sicchè la durata media di esposizione al rischio per gli

Indicando con Δ' la durata media durante la quale risultano esclusi dal rischio i d_t'' eliminati provenienti dagli individui

---

entrati (e di non esposizione per gli usciti), durante il detto intervallo unitario, risulta:

$$\Delta_n = \frac{1}{n} \frac{\sum_{i=0}^{n-1} (n-i)\left(1-\frac{S_t}{n}\right)^i}{\sum_{i=0}^{n-1}\left(1-\frac{S_t}{n}\right)^i}$$

A parte il fattore $\frac{1}{n}$, il denominatore di Δ_n è la somma dei primi n termini di una progressione geometrica il cui primo termine è 1 e la ragione è $1-\frac{S_t}{n}$; il numeratore è la somma delle somme dei termini di n progressioni geometriche aventi tutte come primo termine la unità e come ragione $1-\frac{S_t}{n}$, e aventi n termini la prima, $n-1$ termini la seconda, ... 1 termine solo la n^{esima}.

Risulta quindi:

$$\sum_{i=0}^{n-1}\left(1-\frac{S_t}{n}\right)^i = \frac{1-\left(1-\frac{S_t}{n}\right)^n}{\frac{S_t}{n}},$$

$$\sum_{i=0}^{n-1}(n-i)\left(1-\frac{S_t}{n}\right)^i = \sum_{i=1}^{n}\frac{1-\left(1-\frac{S_t}{n}\right)^i}{\frac{S_t}{n}} = \frac{n-\left(1-\frac{S_t}{n}\right)\frac{1-\left(1-\frac{S_t}{n}\right)^n}{\frac{S_t}{n}}}{\frac{S_t}{n}}$$

da cui

$$\Delta_n = \frac{1}{1-\left(1-\frac{S_t}{n}\right)^n} - \frac{1-\frac{S_t}{n}}{S_t}.$$

E per $n \to \infty$ si ricava da questa

$$\Delta = \lim_{n\to\infty}\Delta_n = \frac{1}{1-e^{-S_t}} - \frac{1}{S_t} = \frac{1}{q_t} - \frac{1}{S_t}.$$

Più rapidamente si può giungere allo stesso risultato, tenendo presente che, dalla definizione di d_t, E_t, M_t, discende

$$\Delta d_t = E_t - M_t$$

da cui

$$\Delta = \frac{E_t}{d_t} - \frac{M_t}{d_t} = \frac{1}{q_t} - \frac{1}{S_t}$$

che si sono aggiunti in seguito per eccesso degli entrati sugli usciti, si trova (1)

$$\Delta' = \frac{1}{S_t}\left[\frac{1}{\dfrac{1}{\dfrac{1}{q_t}-\dfrac{1}{S_t}}-1}-1\right] = \frac{1}{S_t}\,\frac{2\Delta-1}{1-\Delta}$$

Ponendo mente che è

$$d_t' = d_t\,\frac{l_t}{l_t + \Delta K_t}$$

$$d_t'' = d_t\,\frac{\Delta K_t}{l_t\,\Delta K_t}$$

---

(1) Suddiviso, come alla nota precedente, l'intervallo unitario $t\,|\!-\,t+1$ in n intervalli istantanei uguali tra loro, gli entrati nel 1°, 2°, . . . $i+1^{\text{esimo}}$ intervallo risultano proporzionali rispettivamente a $1,\ 1-\frac{S_t}{n}$, ... $\left(1-\frac{S_t}{n}\right)^i$. Poichè gli entrati in ciascun intervallo si riducono, come tutti gli altri esposti al rischio, da un intervallo al successivo, nel rapporto $1-\frac{S_t}{n}$, coll'inizio dell'$i+1^{\text{esimo}}$ intervallo saranno esposti al rischio gli $\left(1-\frac{S_t}{n}\right)^i$ entrati al principio dell'intervallo stesso, gli $\left(1-\frac{S_t}{n}\right)^i$ non eliminati degli $\left(1-\frac{S_t}{n}\right)^{i-1}$ entrati all'inizio dell'intervallo precedente e così via, tutti gli entrati all'inizio degli intervalli dal 1° all'$(i+1)^{\text{esimo}}$ contribuendo per lo stesso numero $\left(1-\frac{S_t}{n}\right)^i$. In complesso, gli esposti al rischio provenienti dagli entrati in osservazione nell'unità di tempo che si considera risultano $(i+1)\left(1-\frac{S_t}{n}\right)^i$, e proporzionale al numero di questi risulta il numero degli eliminati da essi al principio dell'intervallo $(1+i)^{\text{esimo}}$. Dall'istante nel quale entra in osservazione, ciascuno di tali individui resta escluso dalla osservazione per $n-i$ intervalli, e quindi tutti complessivamente restano esclusi dalla osservazione per un tempo $\frac{n-i}{n}(i+1)\left(1-\frac{S_t}{n}\right)^i$. La durata media di non esposizione al rischio risulta quindi per essi :

$$\Delta_n' = \frac{1}{n}\,\frac{\sum_{i=0}^{n-1}(i+1)(n-i)\left(1-\frac{S_t}{n}\right)^i}{\sum_{i=0}^{n-1}(i+1)\left(1-\frac{S_t}{n}\right)^i}$$

A parte il fattore $\frac{1}{n}$, sia il numeratore che il denominatore sono somme di

si può scrivere

$$_M S_t = \frac{d_t}{l_t + \Delta K_t - \Delta'' d_t} \qquad (6\,m)$$

dove è

$$\Delta'' = \frac{\Delta l_t + \Delta' \Delta K_t}{l_t + \Delta K_t}$$

Nella tabella XV sono indicati i valori di Δ e Δ' in funzione di S_t. Col tendere di S_t a 0, Δ tende a 1/2 e Δ' a 1/3; col tendere di S_t all'infinito, Δ e Δ' tendono ad 1.

Determinato il valore approssimato di S_t mediante la (6 m),

---

termini di progressioni geometriche. Tenuto presente che è

$$\sum_{i=0}^{n-1} \left(1 - \frac{S_t}{n}\right)^i = n \frac{1 - \left(\frac{S_t}{n}\right)^n}{S_t}$$

$$\sum_{i=0}^{n-1} i \left(1 - \frac{S_t}{n}\right)^i = n^2 \frac{1 - \left(1 - \frac{S_t}{n}\right)^n}{S_t^2}$$

$$\sum_{i=0}^{n-1} i^2 \left(1 - \frac{S_t}{n}\right)^i = 2n^3 \frac{1 - \left(1 - \frac{S_t}{n}\right)^n}{S_t^3} - n^2 \frac{1 + (2n-3)\left(1 - \frac{S_t}{n}\right)^n}{S_t^2} -$$

$$- n(n-1)^2 \frac{\left(1 - \frac{S_t}{n}\right)^n}{S_t},$$

risulta:

$$\sum_{i=0}^{n-1} (1+i)(n-i)\left(1 - \frac{S_t}{n}\right)^i = \sum_{i=0}^{n-1} \left[n + (n-1)i - i^2\right]\left(1 - \frac{S_t}{n}\right)^i =$$

$$= \frac{n^3}{S_t^3}\left[\left\{1 - \left(1 - \frac{S_t}{n}\right)^n\right\}\frac{S_t^2}{n} + \left\{1 + (1-)\frac{2}{n}\left(1 - \frac{S_t}{n}\right)^n\right\} S_t - \right.$$

$$\left. - 2\left\{1 - \left(1 - \frac{S_t}{n}\right)^n\right\}\right],$$

$$\sum_{i=0}^{n-1} (1+i)\left(1 - \frac{S_t}{n}\right)^i = \frac{n^2}{S_t^2}\left[\left\{\frac{1}{n} - \left(1 - \frac{S_t}{n}\right)^n\right\} S_t + 1 - \left(1 - \frac{S_t}{n}\right)^n\right]$$

e quindi

$$\Delta'_n = \frac{\left\{1 - \left(1 - \frac{S_t}{n}\right)^n\right\}\frac{S_t^2}{n} + \left\{1 + \left(1 - \frac{2}{n}\right)\left(1 - \frac{S_t}{n}\right)^n\right\} S_t - 2\left\{1 - \left(1 - \frac{S_t}{n}\right)^n\right\}}{S_t\left[\left\{\frac{1}{n} - \left(1 - \frac{S_t}{n}\right)^n\right\} S_t + 1 - \left(1 - \frac{S_t}{n}\right)^n\right]}$$

si possono, in base ad esso, determinare i valori di Δ e Δ' e quindi di ${}_M S_t$, in base alla (6 M).

Dal valore di ${}_M S_t$ si ottiene, attraverso la (2 a), il valore di ${}_M q_t$.

Ci si persuade facilmente che, se è $K_t < 0$, è

$$\Delta K_t < \frac{1}{2} K_t, \quad \Delta'' > \Delta > \frac{1}{2}, \quad -\Delta'' d_t < -\frac{1}{2} d_t$$

e quindi ${}_M S_t > {}_m S_t$.

Viceversa, se è $K_t - d_t > 0$ e, a più forte ragione, $K_t > 0$, è

$$\Delta'' < \Delta \text{ e quindi } -\Delta'' d_t > -\Delta d_t,$$

$$\Delta (K_t - d_t) > \frac{1}{2} (K_t - d_t) \text{ e quindi } {}_M S_t < {}_m S_t.$$

Quando è $K_t > 0$, $K_t - d_t < 0$, può essere ${}_M S_t \lesseqgtr {}_m S_t$.

Ci si rende conto facilmente che la differenza tra ${}_M S_t$ e ${}_m S_t$ dipende essezialmente, da una parte, dall'altezza di ${}_m S_t$ e, dall'altra, dall'importanza di K_t in confronto di l_t.

${}_M S_t$ e ${}_M q_t$ indicano dunque il saggio di eliminazione e rispettivamente la probabilità di eliminazione nell'ipotesi che i saggi

---

E per $n \to \infty$ si ricava da questa

$$\Delta' = \lim_{n \to \infty} \Delta'_n = \frac{(1 + e^{-S_t}) S_t - 2 (1 - e^{-S_t})}{S_t (1 - e^{-S_t} - S_t e^{-S_t})}$$

Ricordando che è $1 - e^{-S_t} = q_t$, e quindi $e^{-S_t} = 1 - q_t$, può scriversi anche:

$$\Delta' = \frac{(2 - q_t) S_t - 2 q_t}{S_t [q_t - S_t (1 - q_t)]},$$

dalla quale si ricava facilmente:

$$\Delta' = \frac{1}{S_t} \left[\frac{S_t - q_t}{S_t q_t - (S_t - q_t)} - 1 \right] = \frac{1}{S_t} \left[\frac{1}{\dfrac{1}{\dfrac{1}{q_t} - \dfrac{1}{S_t}} - 1} - 1 \right].$$

Poichè è

$$\frac{1}{q_t} - \frac{1}{S_t} = \Delta,$$

si può scrivere anche:

$$\Delta' = \frac{1}{S_t} \left[\frac{1}{\dfrac{1}{\Delta} - 1} - 1 \right] = \frac{1}{S_t} \frac{2\Delta - 1}{1 - \Delta}.$$

TABELLA XV.

S_t	Δ	Δ'	S_t	Δ	Δ'
10	0.90005	0.8005	0.5	0.5415	0.3620
5	0.8068	0.6348	0.4	0.5332	0.3561
2	0.6565	0.4557	0.3	0.5250	0.3503
1.5	0.6205	0.4236	0.2	0.5167	0.3447
1	0.5820	0.3922	0.1	0.5083	0.3389
0.9	0.5740	0.3861	0.05	0.5042	0.3361
0.8	0.5660	0.3800	0.01	0.5008	0.3339
0.7	0.5579	0.3739	0.005	0.50042	0.33362
0.6	0.5496	0.3679	0.001	0.5000833	0.3335959

istantanei di eliminazione restino costanti durante l'intervallo $t\,|\!-\,t+1$ e che il numero degli entrati e degli usciti (o quanto meno la loro differenza) varii proporzionalmente al numero degli esposti al rischio secondo la legge di eliminazione.

${}_mq_t$, ricavato dalla (5 m) secondo il metodo consueto delle durate medie, fornisce, invece, il valore della probabilità di eliminazione nell'intervallo $t\,|\!-\,t+1$ nell'ipotesi che la probabilità di eliminazione nella frazione $u\,|\!-\,t+1$ dell'intervallo $t\,|\!-\,t+1$ varii proporzionalmente a $t+1-u$ e che gli entrati e gli usciti (o quanto meno le loro differenze) si distribuiscano uniformemente durante l'intervallo. Chiameremo quest'ultimo metodo DMA (*metodo delle durate medie dell'Ackland*) e, in contrapposto ad esso, chiameremo il nostro metodo DMG (*metodo delle durate medie del Gini*).

Il metodo DMG si fonda dunque su ipotesi analoghe per la distribuzione degli eliminati, da una parte, e degli entrati ed usciti, dall'altra, e presenta pertanto una coerenza formale che, dal punto di vista teorico, può essere riguardata come un pregio. Ma ciò non significa che detto metodo sia senz'altro preferibile per le applicazioni pratiche. Converrà esaminare, a tal fine, di caso in caso, se le ipotesi su cui esso si basa rispondano o meno a realtà. Il

metodo potrà, ad esempio, plausibilmente applicarsi quando si tratta di determinare le probabilità annue di morte per le età adulte, per le quali i saggi di eliminazione non variano sostanzialmente entro un anno di vita, mentre può plausibilmente ritenersi che gli entrati e gli usciti siano proporzionali ai viventi. Nel nostro esempio, invece, il metodo DMG non risulta rispondente ; non risponde in particolare a realtà l'ipotesi che i numeri degli entrati variino proporzionalmente agli esposti. Al contrario, il numero delle nuove spose che entrano in osservazione durante un certo anno di età si può ritenere che vada crescendo durante l'anno fino all'età di 20-21 anni, che rappresenta l'età di massima frequenza al matrimonio. L'applicazione che del metodo DMG viene data nelle colonne 5 e 8 della tabella XVI ha quindi solo lo scopo di una esemplificazione fatta per mostrare le differenze tra i valori dei saggi di eliminazione e delle probabilità di eliminazione per concepimento determinati con tale metodo e quelli determinati invece con le altre formule.

Se si confrontano i valori q_t della colonna 6, ottenuti col metodo DEG delle durate esatte, con quelli ${}_M q_t$ della colonna 8, si nota che i secondi sono sempre più bassi dei primi fino a 29 anni compresi. Da 30 anni in su i valori della col. 8 risultano 9 volte inferiori, 6 volte superiori e due volte uguali fino alla 4ª cifra decimale. Rispetto ai valori q_t, i valori ${}_M q_t$ risultano dunque molto meno approssimati che i valori ${}_m q_t$ della colonna 7, ottenuti mediante la formula (6 *m*). Ciò si spiega con la circostanza che, come abbiamo osservato, le ipotesi che stanno a base del metodo DMG non risultano, nel caso in esame, rispondenti a realtà.

* * *

In tutti i ragionamenti precedenti, noi abbiamo considerato le relazioni tra M_t e E_t, e quindi fra S_t e q_t, nell'ipotesi che la riduzione progressiva degli E_t nell'intervallo considerato $t\,|\!-\,t+1$ avvenga esclusivamente per la causa di eliminazione di cui si vuole misurare l'intensità con q_t, ossia, come suol dirsi, secondo la *legge pura di eliminazione* per la causa in esame. La formula (2 *a*) indica dunque la probabilità di eliminazione per una data causa, facendo astrazione dall'influenza che altre cause di eliminazione esercitano sulla riduzione progressiva degli esposti al rischio o, come suol dirsi, indica la *probabilità pura o indipendente di eliminazione* per la causa in esame. Per esempio, noi abbiamo determinato la pro-

Tabella XVI.

Case sovrane, Case mediatizzate tedesche e Case principesche di vari Stati.

Età	q_t Formula (5)	mq_t Formula (5 m)	S_t Formula (6)	mS_t Formula (6 m)	MS_t Formula (6 M)	q_t Formula (2 a) attraverso la (6)	mq_t Formula (2 a) attraverso la (6 m)	Mq_t Formula (2 a) attraverso la (6 M)
	1	2	3	4	5	6	7	8
16–18. . .	0.7356	0.7561	1.1413	1.2157	0.9313	0.6806	0.7035	0.6060
19. . .	0.6957	0.7114	1.0339	1.1042	0.9668	0.6444	0.6685	0.6193
20. . .	0.6208	0.6039	0.8796	0.8652	0.7702	0.5851	0.5790	0.5371
21. . .	0.6091	0.6121	0.8551	0.8821	0.8073	0.5748	0.5861	0.5539
22. . .	0.4609	0.4547	0.5912	0.5886	0.5555	0.4463	0.4449	0.4262
23. . .	0.4146	0.4139	0.5282	0.5219	0.5074	0.4103	0.4066	0.3979
24. . .	0.4201	0.4290	0.5354	0.5462	0.5287	0.4146	0.4209	0.4106
25. . .	0.3577	0.3608	0.4310	0.4403	0.4292	0.3501	0.3562	0.3490
26. . .	0.3397	0.3333	0.4028	0.4000	0.3937	0.3316	0.3297	0.3254
27. . .	0.3108	0.2857	0.3722	0.3333	0.3287	0.3108	0.2834	0.2801
28. . .	0.2472	0.2486	0.2818	0.2838	0.2812	0.2456	0.2471	0.2451
29. . .	0.2444	0.2442	0.2821	0.2782	0.2761	0.2458	0.2429	0.2413
30. . .	0.2030	0.2039	0.2249	0.2270	0.2255	0.2014	0.2031	0.2019
31. . .	0.1926	0.1921	0.2123	0.2125	0.2117	0.1913	0.1914	0.1908
32. . .	0.1562	0.1559	0.1705	0.1691	0.1685	0.1568	0.1556	0.1551
33. . .	0.1078	0.1085	0.1129	0.1147	0.1145	0.1068	0.1084	0.1082
34. . .	0.1339	0.1327	0.1444	0.1421	0.1419	0.1345	0.1325	0.1323
35. . .	0.0841	0.0828	0.0855	0.0864	0.0863	0.0819	0.0828	0.0827
36. . .	0.1124	0.1104	0.1187	0.1169	0.1169	0.1119	0.1103	0.1103
37. . .	0.0768	0.0781	0.0793	0.0813	0.0812	0.0763	0.0781	0.0780
38. . .	0.0707	0.0699	0.0738	0.0725	0.0725	0.0711	0.0699	0.0699
39. . .	0.0539	0.0542	0.0553	0.0557	0.0557	0.0538	0.0542	0.0542
40. . .	0.0234	0.0233	0.0237	0.0235	0.0235	0.0234	0.0232	0.0232
41. . .	0.0382	0.0384	0.0389	0.0392	0.0392	0.0382	0.0384	0.0384
42. . .	0.0127	0.0126	0.0128	0.0127	0.0127	0.0127	0.0126	0.0126
43. . .	0.0133	0.0132	0.0134	0.0133	0.0133	0.0133	0.0132	0.0132
44. . .	0.0057	0.0057	0.0057	0.0057	0.0057	0.0057	0.0057	0.0057
45. . .	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061
46. . .	0.0034	0.0033	0.0034	0.0033	0.0033	0.0034	0.0033	0.0033
16–46. . .	0.1732	0.1721	0.1894	0.1883	0.1875	0.1725	0.1716	0.1710

babilità di un concepimento delle coniugate nell'ipotesi che il numero delle coniugate esposte a detto rischio si riducesse progressivamente solo per i sopravvenuti concepimenti e non anche per morte, vedovanza, divorzio, ecc. Tale procedimento è rispondente quando ci si proponga di misurare la fecondità *naturale* della donna. Chi volesse misurare, invece, la sua fecondità *effettiva*, dovrebbe tener conto anche di queste cause di riduzione delle esposte al rischio. In contrapposto alla legge pura di eliminazione, la successione di eliminati che ne risulta si chiama *legge dipendente di eliminazione*, in quanto essa dipende, oltre che dal concepimento, anche dalle altre cause di eliminazione e corrispondentemente la probabilità di eliminazione, così determinata, si chiama *probabilità dipendente di eliminazione*.

Se noi indichiamo con T_t il saggio di eliminazione globale per tutte le cause che vogliamo prendere in considerazione ai fini della progressiva riduzione degli E_t, otterremo, invece della (2), la

$$M_t' = E_t \frac{1 - e^{-T_t}}{T_t} \qquad (2')$$

e, invece della (2 *a*), la

$$q_t' = S_t' \frac{1 - e^{-T_t}}{T_t} \qquad (2\,a')$$

dove M_t', S_t', q_t' indicano rispettivamente il numero medio degli esposti al rischio, il saggio di eliminazione e la probabilità di eliminazione, in base alla legge dipendente di eliminazione.

Ci si persuade facilmente che è $S_t' = S_t$ perchè il diverso numero degli esposti nei successivi istanti, che si verifica secondo la legge pura e secondo la legge dipendente di eliminazione, non influisce nè sull'altezza dei saggi istantanei di eliminazione (1) nè, per essere questi costanti in tutto l'intervallo $t \,|\!-\, t + 1$, sulla loro media.

Potremo quindi scrivere l'uguaglianza :

$$q_t' = S_t \frac{1 - e^{-T_t}}{T_t}\,, \qquad (2\,a')$$

---

(1) Una dimostrazione della proposizione che i saggi istantanei indipendenti sono uguali ai saggi istantanei dipendenti è stata data da I. Messina, *Le probabilità parziali nella matematica attuariale*, in « Bollettino di Notizie sul Credito e sulla Previdenza » Ministero di Agricoltura, Industria e Commercio, luglio-dicembre 1905, n. 4-6, pag. 42.

la quale ci dà la probabilità di eliminazione per la causa in esame dipendente da tutte le cause di eliminazione considerate.

Per esempio, la (2 *a*) ci ha fornito la probabilità di un concepimento durante un anno di vita nell'ipotesi che nessuna delle donne coniugate morisse o divorziasse o restasse vedova nell'intervallo di età considerato. La (2 *a'*) è atta a fornirci invece la probabilità di un concepimento durante un anno, tenuto conto dell'eliminazione delle coniugate, per morte, vedovanza o divorzio.

Se, infine, volessimo ottenere la probabilità che una coniugata fosse eliminata dalle esposte al rischio del concepimento per concepimento, morte, divorzio o vedovanza, dovremmo usare la formula

$$q_t'' = 1 - e^{-T_t} \qquad (2\ a'')$$

La (2 *a''*) dà la probabilità di eliminazione globale ; la (2 *a*) la probabilità pura o indipendente di eliminazione parziale, la (2 *a'*) la probabilità dipendente di eliminazione parziale.

Vale la pena di rilevare come, nell'ipotesi di saggi istantanei di eliminazione costanti, la relazione tra la probabilità pura e la probabilità dipendente di eliminazione divenga estremamente semplice.

È, come facilmente si intende,

$$q_t'' > q_t > q_t'$$

La tabella seguente mostra la portata delle differenze :

TABELLA XVII.

Case sovrane d'Europa, Case mediatizzate tedesche, e Case principesche di Germania, Austria, Ungheria, Russia, Polonia, Italia, Francia e Gran Bretagna.

ETA'	S_t	T_t	q_t'	q_t	q_t''
–20	1.0711	1.0838	0.6539	0.6574	0.6617
20–24	0.6275	0.6399	0.4635	0.4661	0.4727
25–29	0.3413	0.3557	0.2872	0.2892	0.2993
30–34	0.1715	0.1904	0.1562	0.1576	0.1734
35–39	0.0829	0.1062	0.0786	0.0796	0.1007
40–44	0.0195	0.0401	0.0191	0.0193	0.0393
45–46	0.0048	0.0385	0.0047	0.0048	0.0378
16–46	0.1894	0.2092	0.1709	0.1725	0.1888

Come si vede, le differenze tra i valori di q_t e q_t' sono piccole, mentre sono più forti, e crescenti con l'età, quelle tra i valori di q_t'' e q_t.

I valori di S_t e T_t, anzichè con la formula (6), possono essere determinati, quando ne sia il caso, con le formule (6 m) o (6 M). Tra i valori ${}_mq_t$, ${}_mq_t'$, ${}_mq_t''$, o rispettivamente tra i valori ${}_Mq_t$, ${}_Mq_t'$, ${}_Mq_t''$, che ne risultano, passano relazioni corrispondenti a quelle tra i valori q_t, q_t', q_t'', sopra indicati.

* * *

Le eliminazioni degli esposti al rischio si possono distinguere in *definitive* e *temporanee.*

Le definitive si possono a loro volta distinguere in *definitive in via assoluta,* come sono le eliminazioni per morte, e in *definitive relativamente all'intervallo considerato* $t \mid\!\!- t+1$. Quest'ultimo è, per esempio, il caso per le eliminazioni per concepimento rispetto all'intervallo di un anno di età, in quanto è eccezionale, e ai fini pratici si può trascurare, il caso di una donna che abbia due concepimenti successivi durante lo stesso anno di età.

Temporanee sono invece quelle eliminazioni la cui durata è sempre — o quanto meno può essere, con frequenza non trascurabile — più breve dell'intervallo $t \mid\!\!- t+1$, per modo che si deve tener conto della probabilità che l'eliminato rientri in osservazione durante l'intervallo considerato $t \mid\!\!- t+1$. Tale è il caso nell'eliminazione per malattia rispetto all'intervallo di un anno di età (1).

Conviene distinguere in questo campo il caso A), in cui il gruppo degli individui considerato era esposto al rischio anche prima dell'inizio t dell'intervallo considerato, dal caso B) in cui invece non lo era.

Quest'ultima ipotesi si verifica, quando, ad esempio, si tratti di determinare il quoziente di morbilità nel primo anno di assicurazione o il quoziente di eliminazione per ferite nel primo pe-

---

(1) Il fatto che uno stesso esposto al rischio possa essere eliminato più volte nello stesso intervallo fa sì che solo impropriamente si possa parlare, in tal caso, di *probabilità di eliminazione :* tuttavia, poichè tale espressione è nell'uso corrente (e così correntemente si parla di *probabilità di malattia*), la manterremo, pur tenendo presente il significato particolare che in questo caso essa assume.

riodo della guerra. La prima può verificarsi, invece, quando si tratti di determinare il quoziente di morbilità negli anni di assicurazione successivi o, analogamente, il quoziente di eliminazione per ferite in un secondo periodo della guerra.

Ammesso che il saggio istantaneo di eliminazione e la durata media di eliminazione sieno costanti, il numero probabile degli individui che l'eliminazione temporanea sottrae al gruppo risulta, nel caso *A*), uguale in tutti gli istanti dell'intervallo $t \,|\!-\, t+1$, e quindi il numero probabile E_t degli esposti al rischio all'inizio dell'intervallo $t \,|\!-\, t+1$ è uguale al numero medio M_t degli esposti durante tutto l'intervallo. In questo caso particolare, il numero ε_t degli eliminabili durante l'intervallo $t \,|\!-\, t+1$, non è uguale al numero E_t degli esposti al rischio all'inizio t dell'intervallo, ma è uguale a questo numero aumentato del numero degli individui che all'inizio t restano sottratti al rischio perchè temporaneamente eliminati.

La relazione tra q_t e S_t è in tal caso molto semplice. Il valore probabile di ε_t risulta infatti:

$$\varepsilon_t = E_t + D\,S_t = M_t + D\,S_t\,, \tag{4}$$

dove D indica la durata media dell'eliminazione temporanea, assunta come unità la durata dell'intervallo $t \,|\!-\, t+1$; e quindi è

$$q_t = \frac{d_t}{\varepsilon_t} = S_t \frac{M_t}{M_t + D\,S_t}\,. \tag{4 a}$$

Quando gli elementi della rilevazione rendano consigliabile di ricorrere al metodo delle durate esatte, resta applicabile, per la determinazione di S_t, la formula (6) con l'avvertenza che con r_u si indichino, non solo i nuovi esposti al rischio che sopravvengono in osservazione, ma anche quelli che ritornano dopo il periodo di eliminazione temporanea, mentre, per passare da q_t a S_t, si può far ricorso alla formula seguente, che dalla (4 *a*) immediatamente si deduce:

$$q_t = S_t \frac{\Sigma\,\delta_e}{\Sigma\,\delta_e + \Sigma\,\delta_d}\,, \tag{4 b}$$

dove con $\Sigma\,\delta_e$ si indica la somma delle durate di esposizione al rischio e con $\Sigma\,\delta_d$ si indica la somma delle durate di eliminazione temporanee, assunta sempre come unità la durata dell'intervallo $t \,|\!-\, t+1$.

Ammesso che il saggio istantaneo e la durata media di eliminazione sieno costanti, $\Sigma \delta_d$ può ritenersi uguale alla somma delle durate di eliminazione per i d_t eliminati nell'intervallo $t \,|\!-\, t+1$ (1). Indicando con u l'istante dell'intervallo $t \,|\!-\, t+1$ in cui l'eliminazione temporanea ha inizio e con g il successivo istante in cui essa ha fine, si ha

$$\Sigma \delta_d = \int_t^{t+1} d_u \,(g-u)\, d\,u$$

e quindi

$$q_t = \frac{d_t}{l_t + \int_t^{t+1} (s_u + r_u - d_u - w_u - e_u)\,(t+1-u)\,d\,u + d_u\,(g-u)\,d\,u}, \qquad (7)$$

Questa formula, confrontata con la (6), serve a mettere in luce la relazione tra q_t e S_t.

Più comodo può riuscire, quando se ne abbiano gli elementi, di ricorrere, per la determinazione di q_t, alla formula seguente

$$q_t = \frac{d_t}{l_t + \int_t^{t+1} (s'_u + r'_u - w'_u - e'_u)\,(t+1-u)\,d\,u} \qquad (8)$$

in cui i simboli s'_u, r'_u, w'_u, e'_u hanno gli stessi significati che nella (5), ma sono riferiti, non agli esposti al rischio, ma a tutti i compo-

---

(1) Se con $\Sigma \delta'_d$ si indica la somma delle durate residue di eliminazione degli individui che all'inizio t dell'intervallo erano temporaneamente sottratti al rischio per eliminazione temporanea, con $\Sigma \delta''_d$ la somma delle durate di eliminazione dei d_t individui eliminati temporaneamente nell'intervallo $t \,|\!-\, t+1$ sino al termine $t+1$ dell'intervallo e con $\Sigma \delta'''_d$ la somma delle durate residue di eliminazione degli individui che al termine $t+1$ dell'intervallo sono temporaneamente sottratti al rischio per eliminazione temporanea, sarà

$$\Sigma \delta_d = \Sigma \delta'_d + \Sigma \delta''_d$$

$$\int_t^{t+1} d_u\,(g-u)\,d\,u = \Sigma \delta''_d + \Sigma \delta'''_d .$$

Ora, quando il saggio istantaneo e la durata media di eliminazione sieno costanti, può ammettersi $\Sigma \delta'_d = \Sigma \delta'''_d$ e quindi:

$$\Sigma \delta_d = \int_t^{t+1} d\,u\,(g-u)\,d\,u$$

nenti del gruppo in osservazione, sieno essi esposti al rischio, sieno ad esso temporaneamente sottratti per eliminazione temporanea.

Dal confronto della (8) con la (5) risulta che la formula, la quale fornisce la probabilità di eliminazione per cause transitorie col metodo delle durate esatte per saggi istantanei di eliminazione costanti (metodo DEG), coincide con la formula che fornisce la probabilità di eliminazione per cause definitive col metodo delle durate esatte per probabilità di eliminazione proporzionali (metodo DEA), quando si abbia l'avvertenza di sostituire in questa, agli individui esposti al rischio, tutti gli individui componenti il gruppo in osservazione.

Dalla (4 *b*) e dalla (8) si ricava la seguente formula che, quando si abbiano gli elementi per applicarla, può riuscire più comoda per la determinazione di S_t:

$$S_t = \frac{d_t}{l_t + \int_t^{t+1} (s'_u + r'_u - w'_u - e'_u)\,(t + 1 - u)\,d\,u - \Sigma\,\delta_d} \qquad (9)$$

La determinazione di q_t in base alla (8) e di S_t in base alla (9) è stata eseguita sui dati relativi alla morbilità per gruppi quinquennali di età degli assicurati alla Cassa Nazionale Malattie per gli Addetti al Commercio nel triennio 1930-32 per una provincia ad alta e per una provincia a bassa morbilità (1). I risultati sono contenuti nelle colonne 5 e 6 delle tabelle XVIII e XIX.

Si noti che, nel caso della eliminazione per cause transitorie, non è necessario passare attraverso il valore di S_t al fine di giungere al valore di q_t in base all'ipotesi dei saggi istantanei di eliminazione costanti.

Si noti pure che, in detto caso, il numero degli esposti al rischio, a parte cause estranee di eliminazione, resta costante, cosicchè resta costante anche il numero degli entrati e quello degli usciti nella ipotesi che essi sieno proporzionali al numero degli esposti al rischio. Questa ipotesi viene dunque, in detto caso, a coincidere con l'ipotesi della distribuzione uniforme degli entrati e degli usciti. La durata media, durante la quale gli entrati restano in osservazione e gli usciti ne sono esclusi, è, ovviamente,

(1) I dati furono cortesemente comunicati dalla Cassa ; le elaborazioni furono eseguite dall'Ing. Vincenzo Curatolo, studente della Scuola di Scienze Statistiche e Attuariali, sotto la sorveglianza del Dr. V. Castellano.

in tale ipotesi, da porre $= \frac{1}{2}$, così che le formule per la determinazione della probabilità di eliminazione e del saggio medio di eliminazione, in base al metodo delle durate medie, divengono:

$$ {}_m q_t = \frac{d_t}{l_t + \frac{1}{2} K'} \qquad (8\,m) $$

$$ {}_m S_t = \frac{d_t}{l_t + \frac{1}{2} K' - \Sigma\, \delta_u} \qquad (9\,m) $$

dove è

$$ K' = \Sigma\, s'_u + r'_u - w'_u - e'_u $$

Il confronto tra i valori di q_t della col. 6, ottenuti con la (8), e i valori di ${}_m q_t$ della col. 7, ottenuti con la (8 m), mostrano come le approssimazioni non sieno cattive, tenuto conto del numero non molto ampio delle osservazioni, a cui si riferiscono i dati della tabella XIX (cfr. col. 2, 3, 4). Nella tabella XVIII, i valori di ${}_m q_t$ risultano 3 volte inferiori, 2 volte uguali fino alla 4 a decimale, e 5 volte superiori ai valori di q_t; nella tabella XIX, 4 volte inferiori e 6 volte superiori.

* * *

Ritorniamo all'ipotesi B), in cui si suppone che gli individui considerati non fossero esposti al rischio prima dell'inizio dell'intervallo $t \,|\!-\, t+1$, per modo che, all'inizio t, tutti erano esposti al rischio e nessuno ne era, sia pure temporaneamente, sottratto. In tal caso, il numero degli eliminabili ε_t durante l'intervallo $t \,|\!-\, t+1$ coincide col numero E_t degli esposti al rischio all'inizio dell'intervallo, ma il numero probabile degli eliminati non resta costante nei successivi istanti dell'intervallo, ma cresce, più rapidamente prima, meno rapidamente poi, tendendo asintoticamente ad un limite che è rappresentato dal numero degli eliminati che si verificherebbe, con gli stessi saggi istantanei di eliminazione, nell'ipotesi A).

Nell'ipotesi B) la relazione tra il numero E_t degli esposti al rischio all'inizio dell'intervallo e il numero medio M_t degli esposti al rischio durante l'intervallo, e, corrispondentemente, la relazione rigorosa tra q_t e S_t, vengono espresse da formule meno semplici.

* * *

Un gruppo di individui, come può essere esposto contemporaneamente a parecchie cause di eliminazione definitiva, così può essere esposto contemporaneamente a parecchie cause di eliminazione temporanea (per es. malattie e ferite) e può anche essere esposto, ad un tempo, ad una od alcune cause di eliminazione temporanea (per es. malattie e ferite), da una parte, e ad una od alcune cause di eliminazione definitiva (per es. morte e prigionia), dall'altra. Anche in casi siffatti può interessare di determinare la probabilità di eliminazione globale o la probabilità pura di eliminazione parziale o la probabilità dipendente di eliminazione parziale.

Soffermiamoci sul caso dell'intervento di una causa di eliminazione temporanea che talvolta determina l'intervento di una causa di eliminazione definitiva, come avviene per le malattie e le ferite che talvolta determinano la morte. In tal caso appare naturale che, nel determinare la probabilità di eliminazione, si tenga conto della progressiva riduzione degli individui in osservazione per effetto della causa definitiva di eliminazione (che spesso della causa temporanea considerata costituisce una conseguenza), determinando così la probabilità di eliminazione parziale per la causa di eliminazione temporanea dipendente anche dalla causa di eliminazione definitiva.

Se q_t è la probabilità di eliminazione parziale per la causa di eliminazione temporanea determinata secondo la (4 *a*) o la (7) o la (8), la probabilità di eliminazione parziale dipendente anche dalla causa di eliminazione definitiva sarà data dalla:

$$q_t' = q_t \frac{1 - e^{-\sigma_t}}{\sigma_t} \qquad (4'\,a)$$

dove σ_t è il saggio medio di eliminazione dipendente dei soggetti esposti al rischio nell'intervallo $t \,|\!-\, t+1$ quale può ricavarsi dalla (6).

Si osservi infatti che, in tal caso, sono esposti al rischio di eliminazione dipendente tutti gli individui in osservazione, sieno essi esposti alla causa di eliminazione temporanea, o sieno invece momentaneamente eliminati per detta causa temporanea. Indicando con M_t' il numero medio degli esposti al rischio di eliminazione definitiva, sarà, tenendo presente la (4):

$$M_t' = M_t + D\,S_t = \varepsilon_t\,.$$

Ne segue

$$q_t = \frac{d_t}{\varepsilon_t} = \frac{d_t}{M_t'} = S_t'\,.$$

Da questa uguaglianza e dalla (2 *a'*), si ottiene appunto la (4' *a*).

Nelle colonne 3 e 4 delle tabelle XVIII e XIX, sono dati i valori del saggio medio di mortalità σ_t e dell'espressione $\frac{1 - e^{-\sigma_t}}{\sigma_t}$ per gli assicurati della Cassa Nazionale Malattie per gli addetti al Commercio, e, nella colonna 5, i corrispondenti valori della probabilità dipendente di malattia q_t', che possono porsi a confronto coi rispettivi valori della probabilità pura q_t indicati alla colonna 2. Le differenze fra q_t e q_t' sono dell'ordine di qualche millesimo per le età più basse e vanno crescendo con l'età fino a raggiungere l'ordine di qualche centesimo per le età più elevate.

TABELLA XVIII.

Provincia ad alta morbilità.

ETÀ	l_t	d_t	K_t'	S_t	q_t	$_m q_t$	σ_t	$\frac{1-e^{-\sigma_t}}{\sigma_t}$	q_t'
1	2	3	4	5	6	7	8	9	10
15-19	2556	522	754	0.1765	0.1741	0.1780	0 00467	0.9979	0.1737
20-24	3321	721	−28	0.2216	0.2178	0.2180	0.00483	0.9979	0.2173
25-29	3441	822	−62	0.2480	0.2430	0.2411	0.00532	0.9962	0.2421
30-34	2794	725	−212	0.2761	0.2697	0.2697	0.00893	0.9955	0.2685
35-39	2069	509	−107	0.2590	0.2533	0.2525	0.01394	0.9928	0.2515
40-44	1646	373	−50	0.2341	0.2294	0.2301	0.02091	0.9895	0.2270
45-49	1223	232	−61	0.1957	0.1923	0.1945	0.02776	0.9863	0.1765
50-54	869	180	−81	0.2220	0.2173	0.2137	0.03985	0.9804	0.2130
55-59	638	115	−51	0.1953	0.1915	0.1878	0.04663	0.9771	0.1871
60-64	386	85	−36	0.2375	0.2325	0.2310	0.04650	0.9772	0.2272
15-64	18943	4284	66	0.2297	0.2253	0.2258	0.01298	0.9938	0.2228

TABELLA XIX.

Provincia a bassa morbilità.

ETA'	l_t	d_t	K'_t	S_t	q_t	mq_t	σ_t	$\frac{1-e^{-\sigma_t}}{\sigma_t}$	q'_t
1	2	3	4	5	6	7	8	9	10
15–19	829	68	252	0.0688	0.0685	0.0712	0.00907	0.9956	0.0682
20–24	1022	97	−32	0.0975	0.0970	0.0964	0.00700	0.9971	0.0967
25–29	1036	124	−66	0.1225	0.1215	0.1236	0.00980	0.9949	0.1209
30–34	738	99	−46	0.1368	0.1358	0.1385	0.01372	0.9934	0.1349
35–39	584	95	10	0.1586	0.1571	0.1613	0.02810	0.9861	0.1549
40–44	495	82	−14	0.1696	0.1679	0.1680	0.04095	0.9797	0.1645
45–49	444	64	−19	0.1502	0.1483	0.1473	0.06257	0.9693	0.1437
50–54	355	37	−20	0.1090	0.1078	0.1072	0.06704	0.9672	0.1043
55–59	220	24	−21	0.1168	0.1152	0.1146	0.08161	0.9603	0.1106
60–64	129	12	−17	0.1003	0.0993	0.0996	0.08271	0.9595	0.0953
15–64	5852	702	27	0.1191	0.1182	0.1197	0.02526	0.9873	0.1167

RIASSUNTO

In questa memoria abbiamo trattato della determinazione dei quozienti di eliminazione (probabilità di eliminazione, saggio medio di eliminazione) durante un intervallo di esistenza $t \mid\!\!- t+1$, fondandoci sull'ipotesi che i saggi istantanei di eliminazione restino costanti durante tutto l'intervallo considerato.

Generalmente, quando si parla di eliminazione, si intende riferirsi alle cause di eliminazione definitive, e queste noi abbiamo considerate per prime.

Le ipotesi di carattere generale, che comunemente si fanno ai fini della determinazione dei quozienti di eliminazione, sono le seguenti:

a) che la distribuzione degli eliminati sia uniforme nei successivi istanti dell'intervallo $t \mid\!\!- t+1$, ciò che equivale ad ammettere che i saggi istantanei di eliminazione vadano aumentando durante l'intervallo $t \mid\!\!- t+1$. È l'ipotesi, in base a cui generalmente si determina il coefficiente di mortalità quando gli esposti

al rischio restano in osservazione durante tutto l'intervallo $t\,|\!-\,t+1$ e sulla quale è basata la nota formula di Becker ;

b) che la probabilità di eliminazione dall'istante u, interno all'intervallo $t\,|\!-\,t+1$, alla fine $t+1$ di detto intervallo sia proporzionale alla frazione di intervallo $t+1-u$, ciò che equivale ad ammettere che i saggi istantanei di eliminazione vadano diminuendo durante l'intervallo $t\,|\!-\,t+1$. È l'ipotesi che sta a base della determinazione della probabilità di eliminazione col metodo delle durate esatte, proposto dall'Ackland ed applicato quando gli esposti al rischio non restano in osservazione durante tutto l'intervallo $t\,|\!-\,t+1$, ma si verificano, nel gruppo osservato, movimenti di entrata e movimenti di uscita all'infuori della causa di eliminazione considerata.

È chiaro che sia l'una che l'altra di tali ipotesi non può corrispondere sempre a realtà e che talvolta nè l'una nè l'altra è ammissibile neppure in via di prima approssimazione.

Non di rado si avvicina maggiormente a realtà l'ipotesi che i saggi istantanei di eliminazione siano costanti durante l'intervallo $t\,|\!-\,t+1$; e, d'altra parte, è a questa ipotesi che conviene ricorrere, in base al principio di indifferenza, quando nulla si conosca circa l'andamento dei saggi istantanei di eliminazione durante il detto intervallo. Eppure, che io sappia, tale ipotesi, non venne finora presa a base della determinazione della probabilità di eliminazione.

Fu, invece, considerata l'ipotesi che il numero dei morti cresca o diminuisca in funzione lineare del tempo durante l'intervallo $t\,|\!-\,t+1$; ma le formule che se ne sono ricavate non sono di facile applicazione, cosicchè non vennero adottate nella pratica.

In questa memoria noi abbiamo anzitutto determinato le formule corrispondenti a quella di Becker e a quella del metodo delle durate esatte di Ackland nell'ipotesi che i saggi istantanei di eliminazione sieno costanti nell'intervallo $t\,|\!-\,t+1$.

A parità di saggio medio di eliminazione, la formula di Becker conduce a valori della probabilità di eliminazione q_t praticamente uguali a quelli a cui si giunge con la formula nostra quando il saggio medio di eliminazione sia uguale o minore di 0,1, mentre, per valori di questo maggiori di 0.1, essa conduce a valori di q_t sensibilmente superiori e tanto più superiori quanto più il valore del saggio medio di eliminazione è elevato. Quando il saggio medio di

eliminazione sia molto elevato (maggiore di 2), la formula di Becker non si può più adoperare perchè porterebbe a valori di $q_t > 1$.

La determinazione di q_t con la formula delle durate esatte di Ackland (metodo DEA) conduce, nelle applicazioni da noi eseguitene, a valori praticamente uguali a quelli ottenuti con la formula delle durate esatte da noi proposta (metodo DEG) per valori di $q_t \lesssim 0.15$, mentre, per valori di $q_t > 0.15$, essa conduce a valori sensibilmente diversi e generalmente superiori. Per valori di q_t molto elevati, le differenze riscontrate hanno, in qualche caso, oltrepassato il 10 %. Abbiamo messo in evidenza come anche il metodo DEA possa portare a valori di $q_t > 1$. Ciò può avvenire quando la probabilità è molto elevata e molti esposti al rischio entrano in osservazione nell'ultima parte dell'intervallo $t \,|\!-\, t+1$.

In pratica gli attuari, anzichè al metodo delle durate esatte, ricorrono ai metodi, pure proposti dall'Ackland, delle durate medie (metodo DMA) o delle durate più vicine, i quali del metodo delle durate esatte costituiscono delle semplificazioni e si basano sulla stessa ipotesi del metodo delle durate esatte per ciò che concerne l'andamento delle probabilità di eliminazione durante l'intervallo $t \,|\!-\, t+1$ e in più sulla ipotesi che la distribuzione degli entrati e degli usciti durante detto intervallo sia uniforme.

Noi abbiamo indicato le formule corrispondenti a cui si perviene, ammettendo, invece, da una parte, che i saggi istantanei di eliminazione restino costanti durante l'intervallo $t \,|\!-\, t+1$ e, dall'altra : *a*) che si distribuiscano uniformemente nell'intervallo le differenze tra entrati ed usciti o eliminati ; oppure : *b*) che resti costante durante l'intervallo il rapporto della differenza tra entrati ed usciti agli esposti al rischio. Quest'ultimo metodo (metodo DMG) sembra particolarmente rispondente per la determinazione della mortalità nelle età adulte.

Oltre che della eliminazione globale, che è dovuta al complesso di tutte le cause di eliminazione considerate, abbiamo trattato della eliminazione parziale per una o per alcune di dette cause, considerando a parte la probabilità pura o indipendente di eliminazione parziale e la probabilità di eliminazione parziale dipendente anche dalle altre cause di eliminazione. Per ciascun caso, abbiamo dato le formule appropriate nell'ipotesi di saggi istantanei di eliminazione costanti. In tale ipotesi, le relazioni tra probabilità pura e probabilità dipendente risultano molto semplici.

All'esame delle cause di eliminazione definitive, abbiamo

fatto seguire l'esame delle cause di eliminazione temporanee, proponendo ed applicando, anche per queste, formule appropriate. Per la probabilità di eliminazione temporanea, abbiamo particolarmente considerato, oltre che le probabilità pure di eliminazione, le probabilità di eliminazione dipendenti anche da cause di eliminazione a carattere definitivo.

Se non ci inganniamo, i metodi qui proposti potrebbero essere utilmente applicati nel campo delle discipline statistiche e in particolare in quello delle scienze attuariali. Mediante un esempio, abbiamo mostrato quanto sia grossolano, in confronto ai detti metodi, il metodo dei coefficienti specifici di fecondità matrimoniale, usato dai demografi per misurare la fecondità delle coniugate, e come esso porti ad un'imagine fondamentalmente errata della curva della fecondità matrimoniale della donna secondo l'età.

RÉSUMÉ

Dans ce mémoire, nous avons traité la détermination des quotients d'élimination (probabilité d'élimination, taux moyen d'élimination) pendant un intervalle d'existence $t \,|\!-\, t+1$, en nous basant sur l'hypothèse que les taux instantanés d'élimination restent constants pendant tout l'intervalle considéré.

Généralement, lorsque l'on parle d'élimination, l'on entend se référer aux causes d'élimination définitive et ce sont celles-là que nous avons considérées en premier lieu.

Les hypothèses de caractère général que l'on fait communément aux fins de la détermination des quotients d'élimination, sont les suivantes :

a) que la distribution des éliminés est uniforme dans les instants successifs de l'intervalle $t \,|\!-\!-\, t+1$, ce qui équivaut à admettre que les taux instantanés d'élimination vont en augmentant pendant l'intervalle $t \,|\!-\, t+1$. C'est l'hypothèse sur la base de laquelle l'on détermine généralement le coefficient de mortalité quand les exposés au risque restent en observation pendant tout l'intervalle $t \,|\!-\, t+1$ et sur laquelle est fondée la formule bien connue de Becker.

b) que la probabilité d'élimination de l'instant u, interne à l'intervalle $t \,|\!-\, t+1$, à la fin dudit intervalle $t+1$ est proportionnelle à la fraction d'intervalle $t+1-u$, ce qui équivaut

à admettre que les taux instantanés d'élimination vont en diminuant pendant l'intervalle $t \mid\!\!- t+1$. C'est l'hypothèse qui se trouve à la base de la détermination de la probabilité d'élimination par la méthode des durées exactes, proposée par Ackland et appliquée lorsque les exposés au risque ne restent pas en observation pendant tout l'intervalle $t \mid\!\!- t+1$, mais que se produisent dans le groupe observé des mouvements d'entrée et des mouvements de sortie en dehors de la cause d'élimination considérée.

Il est clair que ni l'une ni l'autre de ces hypothèses ne peut toujours correspondre à la réalité et que parfois, ni l'une ni l'autre n'est admissible même à titre de première approximation. Il n'est pas rare que l'hypothèse que les taux instantanés d'élimination sont constants pendant l'intervalle $t \mid\!\!- t+1$ s'approche davantage à la réalité ; et, d'autre part, c'est à cette hypothèse qu'il convient de recourir, sur la base du principe d'indifférence, lorsque l'on ne sait rien de l'évolution des taux instantanés d'élimination pendant ledit intervalle.

Et cependant, à ma connaissance, cette hypothèse n'a pas été jusqu'ici prise comme base de la détermination de la probabilité d'élimination.

On a considéré par contre l'hypothèse que le nombre des morts croit ou diminue en fonction linéaire du temps pendant l'intervalle $t \mid\!\!- t+1$; mais les formules que l'on en a tirées ne sont pas d'application facile, de telle sorte qu'elles n'ont pas été appliquées dans la pratique.

Dans ce mémoire, nous avons avant tout déterminé les formules correspondant à celle de Becker et à celle de la méthode des durées exactes de Ackland, dans l'hypothèse que les taux instantanés d'élimination sont constants dans l'intervalle $t \mid\!\!- t+1$.

A parité de taux moyen d'élimination, la formule de Becker conduit à des valeurs de la probabilité d'élimination pratiquement égales à celles auxquelles l'on arrive par notre formule, lorsque le taux moyen d'élimination est égal ou inférieur à 0,1, tandis que, pour des valeurs de celui-ci supérieures à 0,1, elle conduit à des valeurs de q_t sensiblement supérieures, et d'autant plus supérieures que la valeur du taux moyen d'élimination est plus élevée. Lorsque le taux moyen d'élimination est très élevé (supérieur à 2), la formule de Becker ne peut plus être adoptée parce qu'elle porterait à des valeurs de $q_t > 1$.

La détermination de q_t par la formule des durées exactes

de Ackland (méthode DEA) conduit, dans les applications faites par nous à des valeurs pratiquement égales à celles obtenues par la formule des durées exactes proposée par nous (méthode DEG) pour des valeurs de $q_t \lesssim 0{,}15$ tandis que, pour des valeurs $q_t > 0{,}15$, elle conduit à des valeurs sensiblement différentes et généralement supérieures. Pour des valeurs de q_t très élevées, les différences constatées ont, dans certains cas, dépassé 10 %. Nous avons mis en évidence que la méthode DEA également peut porter à des valeurs $q_t > 1$. Cela peut se produire lorsque la probabilité d'élimination est très élevée et que de nombreux exposés au risque entrent en observation pendant la dernière partie de l'intervalle.

En pratique, les actuaires, plutôt qu'à la méthode des durées exactes, recourent aux méthodes, proposées également par Ackland, des durées moyennes (méthode DMA) ou des durées les plus voisines, qui constituent des simplifications de la méthode des durées exactes et se basent sur la même hypothèse que la méthode des durées exactes en ce qui concerne l'évolution des probabilités d'élimination pendant l'intervalle $t \mid\!\!- t+1$ et, en outre, sur l'hypothèse que la distribution des entrés et des sortis pendant ledit intervalle est uniforme.

Nous avons indiqué les formules correspondantes auxquelles l'on arrive en admettant, au lieu de cela, d'une part, que les taux instantanés d'élimination restent constants pendant l'intervalle $t \mid\!\!- t+1$ et, de l'autre : *a*) que les différences entre les entrés et les sortis ou éliminés se distribuent uniformément ; ou, *b*) que le rapport de la différence entre les entrés et les sortis avec les exposés au risque reste constant. Cette dernière méthode (méthode DMG) semble particulièrement adaptée pour la détermination de la mortalité à l'âge adulte.

En outre de l'élimination globale qui est due à l'ensemble de toutes les causes d'élimination considérées, nous avons traité l'élimination partielle pour l'une ou pour certaines desdites causes en considérant à part la probabilité pure ou indépendante d'élimination partielle et la probabilité d'élimination partielle dépendant également des autres causes d'élimination. Pour chaque cas, nous avons donné la formule appropriée dans l'hypothèse des taux instantanés d'élimination constants. Dans cette hypothèse, les rapports entre probabilité pure et probabilité dépendante sont très simples.

A l'examen des causes d'élimination définitive, nous avons

fait suivre l'examen des causes d'élimination temporaire, en proposant et en appliquant, pour celles-ci aussi, des formules appropriées.

Pour la probabilité d'élimination temporaire, nous avons particulièrement pris en considération, en outre des probabilités, pures d'élimination, les probabilités d'élimination dépendant également de causes d'élimination à caractère définitif.

Si nous ne nous trompons pas, les méthodes proposées ici pourraient être utilement appliquées dans le domaine des sciences statistiques et en particulier dans celui des sciences actuairielles. Au moyen d'un exemple, nous avons montré combien est grossière, vis-à-vis desdites méthodes, la méthode des coefficients spécifiques de fécondité matrimoniale, employée par les démographes pour mesurer la fécondité des femmes mariées et comment elle porte à une image fondamentalement erronée de la courbe de la fécondité matrimoniale de la femme selon l'âge.

SUMMARY

In this article we have discussed the determination of the quotients of elimination (probability of elimination, average rate of elimination) during a time interval $t \mid\!- t + 1$, on the hypothesis that the rates of elimination at any given instant remain constant during the whole of the time interval considered.

As a rule, in speaking of elimination, it is intended to refer to the definitive causes of elimination and these we have considered in the first instance.

The hypotheses of a general character normally adopted for the purposes of the determination of the quotients of elimination are the following :—

(*a*) that the distribution of the eliminated subjects is uniform in the successive instants of the interval $t \mid\!- t + 1$, which amounts to the admission that the rates of elimination at any given instant increase during the interval $t \mid\!- t + 1$. It is on this hypothesis that the determination of the coefficient of mortality is generally based, when the subjects exposed to the risk remain under observation for the whole period $t \mid\!- t + 1$, and the well known Becker formula depends on this assumption ;

(*b*) that the probability of elimination from the instant u,

within the interval $t \mid\!\!- t + 1$, at the end $t + 1$ of the said interval, is proportional to the fraction of the interval $t + 1 - u$, which is equivalent to saying that the rates of elimination at any given instant diminish during the interval $t \mid\!\!- t + 1$. On this hypothesis is based the determination of the probabilities of elimination by the exact duration method, as recommended by Ackland and applied when the subjects liable to the risk do not remain under observation during the whole of the interval $t \mid\!\!- t + 1$ but movements of entry and movements of withdrawal are to be noted in the group under observation, extraneous to the cause of elimination under consideration.

It is evident that it is impossible for either of these hypotheses always to correspond with the facts, and that sometimes neither the one nor the other can be admitted, even by way of primary approximation.

Not infrequently the hypothesis that the rates of elimination at any given instant remain constant during the interval $t \mid\!\!- t + 1$ represents a closer approximation to reality ; furthermore, on the principle of indifference, it is expedient to employ this hypothesis, when nothing is known of the course of the rates of elimination at any given instant during this interval.

So far as I am aware, however, this hypothesis has not, up to the present, been adopted as the basis for the determination of the probability of elimination.

On the other hand, the hypothesis that the number of deaths increases or diminishes during the interval $t \mid\!\!- t + 1$ in linear function of the time has been considered ; the formulae derived therefrom, however, are difficult to apply and in practice they have not been utilised.

In this article we have in the first place determined the formulae corresponding to Becker's formula and to that of Ackland's exact duration method, on the assumption that the rates of elimination at any given instant are constant in the interval $t \mid\!\!- t + 1$.

The average rate of elimination being equal, Becker's formula gives values for the probability of elimination q_t practically equal to those obtained by our formula, when the average rate of elimination is equal to or less than 0.1, while for values of such rate in excess of 0.1, it gives values for q_t, which are perceptibly higher and increasingly so with any raising of this average rate.

When the average rate is very high (in excess of 2), the Becker formula can no longer be applied, as it would give values for $q_t > 1$.

The determination of q_t by Ackland's formula of exact duration (method DEA) leads in the practical applications made in this article to values which are virtually equal to those obtained by the formula of exact duration suggested by ourselves (method DEG) for values of $q_t \lesssim 0.15$, while for values of $q_t > 0.15$ it leads to values perceptibly diverse and generally higher. For very high values of q_t the differences found have in certain cases exceeded 10 per cent. We have also shown that the method DEA may also lead to values of $q_t > 1$. This may arise when the probability of elimination is very high and many subjects liable to the risk come under observation at the last stage of the interval $t \mid\!— t + 1$.

Professional actuaries in practice make use, in place of the exact duration method, of the mean duration method (method DMA), or of the nearest duration method, both suggested by Ackland. These methods are in effect simplifications of the exact duration method, and depend on the same hypothesis as regards the course of the probabilities of elimination during the interval $t \mid\!— t + 1$, and still more on the hypothesis that the distribution of entries and of withdrawals is uniform during the interval considered.

We have indicated the corresponding formulae, based on the assumption, on the one hand, that the rates of elimination at any given instant remain constant during the interval $t \mid\!— t + 1$, and, on the other :—

(*a*) that the differences between entries and withdrawals or otherwise eliminated subjects are evenly spread over the interval ; or :

(*b*) that the ratio of the difference between entries and withdrawals to the number of subjects liable to the risk remains constant during the period. This last method (method DMG) seems particularly valuable for the determination of mortality rates in the case of adults.

In addition to aggregate elimination, which is due to the combination of all the causes of elimination discussed, we have treated also partial elimination, due to one or more of these causes, dealing separately with the pure independent probability of partial elimination and with the probability of partial elimination contingent on the other causes of elimination. For each case we have

given the appropriate formulae, based on the supposition of rates of elimination constant at any given instant. On this assumption the relations between pure and contingent probability are very simple.

Our examination of the definitive causes of elimination is followed by a discussion of the temporary causes of elimination, for which suitable formulae have also been put forward and applied. Regarding the probability of temporary elimination, we have given special consideration to the probabilities of elimination contingent on definitive causes in addition to the pure probabilities thereof.

If we are not mistaken, the methods here proposed might be also applied with profit in the field of statistical research, and specially in that of the actuarial sciences. By means of a concrete example, we have demonstrated how rough and ready, in contrast with the methods here discussed, is that of specific coefficients for matrimonial fertility, as employed by writers on population questions in assessing the fertility of married women, and how this method leads to a fundamentally false idea of the curve of matrimonial fertility of the woman in relation to her age.

AUSZUG

In dieser Abhandlung haben wir die Bestimmung der Ausscheidequotienten (Ausscheidewahrscheinlichkeit, mittlere Ausscheideziffer) während eines Zeitabschnittes des Bestehens $t\,|\!-\,t+1$ behandelt, wobei wir uns auf die Annahme gründeten, dass die Ausscheideziffern der einzelnen Momente während des ganzen in Betracht gezogenen Zeitabschnittes konstant bleiben.

Allgemein versteht man, wenn man von Ausscheidung spricht, eine Bezugnahme auf die definitiven Ausscheidegründe, und diese haben wir hier auch zunächst betrachtet.

Die Annahmen allgemeiner Art, die gewöhnlich zur Bestimmung der Ausscheidequotienten gemacht werden, sind folgende:

a) dass die Verteilung der Ausgeschiedenen gleichmässig sei in den aufeinanderfolgenden Momenten des Zeitabschnittes $t\,|\!-\,t+1$, was gleichbedeutend ist mit der Annahme, dass die Ausscheideziffern der einzelnen Momente während des Zeitabschnittes $t\,|\!-\,t+1$ wachsen. Es ist dies die Annahme, auf Grund deren man allgemein den Sterbenskoeffizienten bestimmt, wenn

die dem Risiko ausgesetzten Personen während des ganzen Zeitabschnittes $t \mid\!- t+1$ unter Beobachtung bleiben, und auf die sich auch die bekannte Becker'sche Formel gründet;

b) dass die Ausscheidewahrscheinlichkeit vom Moment u innerhalb des Zeitabschnittes $t \mid\!- t+1$ bis zum Ende $t+1$ eben dieses Zeitraumes proportional sei dem Bruchteil des Abschnittes $t+1-u$, was gleichbedeutend ist mit der Annahme, dass die Ausscheideziffern der einzelnen Momente während des Zeitraumes $t \mid\!- t+1$ abnehmen. Es ist dies die Annahme, welche die Grundlage bildet für die Bestimmung der Ausscheidewahrscheinlichkeit nach der Methode der genauen Zeitdauern, die von Ackland vorgeschlagen wurde und zur Anwendung kommt, wenn die dem Risiko ausgesetzten Personen nicht während des ganzen Zeitabschnittes $t \mid\!- t+1$ unter Beobachtung bleiben, sondern sich in der beobachteten Gruppe Eintritts- und Austrittsbewegungen geltend machen, die ausserhalb des in Betracht gezogenen Ausscheidegrundes liegen.

Es ist klar, dass sowohl die eine wie die andere Annahme nicht immer der Wirklichkeit entsprechen kann und dass bisweilen weder die eine noch die andere als erster Näherungswert zulässig ist. Nicht selten nähert sich mehr der Wirklichkeit die Annahme, dass die Ausscheideziffern der jeweiligen Momente während des Zeitabschnittes $t \mid\!- t+1$ konstant seien; und andererseits ist es zweckmässig, nach dem Grundsatz der Indifferenz, Zuflucht zu dieser Annahme zu nehmen, wenn nichts bekannt ist über die Entwicklung der Ausscheideziffern der einzelnen Momente während des genannten Zeitraumes.

Gleichwohl wurde, soviel ich weiss, bis jetzt diese Annahme nicht zur Grundlage für die Bestimmung der Ausscheidewahrscheinlichkeiten genommen.

Es wurde vielmehr die Annahme in Betracht gezogen, dass die Zahl der Todesfälle wachse oder abnehme in linearer Funktion der Zeit während des Zeitabschnittes $t \mid\!- t+1$; die Formeln jedoch, die daraus abgeleitet wurden, sind nicht leicht anwendbar, sodass sie in der Praxis keine Verwendung fanden.

In dieser Abhandlung haben wir vor allem die Formeln bestimmt, die derjenigen von Becker und derjenigen der Ackland'schen Methode der genauen Zeitdauern entsprechen, mit der Annahme, dass die Ausscheideziffern der einzelnen Momente in dem Zeitabschnitt $t \mid\!- t+1$ konstant sind.

Bei gleicher mittlerer Ausscheideziffer führt die Becker'sche Formel zu Werten der Ausscheidewahrscheinlichkeit q_t, die praktisch dejenigen gleich sind, zu denen man mit unserer Formel gelangt, sofern die mittlere Ausscheideziffer gleich oder kleiner ist als 0,1, während sie für Werte derselben, die 0,1 übersteigen, zu Werten von q_t führt, die bedeutend höher liegen und umso höher werden, je mehr der Wert der mittleren Ausscheideziffer sich erhöht. Liegt der Wert der mittleren Ausscheideziffer sehr hoch (über 2), so lässt sich die Becker'sche Formel nicht mehr anwenden, da sie zu Werten $q_t > 1$ führen würde.

Die Bestimmung von q_t mittels der Formel der genauen Zeitdauern von Ackland (Methode DEA), führt in den von uns durchgeführten Anwendungen zu Werten, die praktisch denen gleichkommen, welche mit der von uns vorgeschlagenen Formel der genauen Zeitdauern (Methode DEG) für Werte von $q_t \leqq 0{,}15$ erreicht wurden, während sie für Werte von $q_t > 0{,}15$ zu merklich davon verschiedenen und gewöhnlich höheren Werten führt. Für sehr hohe Werte von q_t haben die festgestellten Differenzen in einigen Fällen über 10 % betragen. Wir haben hervorgehoben, wie auch die Methode DEA zu Werten von $q_t > 1$ führen kann. Das kann vorkommen, wenn die Ausscheidewahrscheinlichkeit sehr hoch ist und viele dem Risiko ausgesetzte Personen im letzten Teil des Zeitabschnittes $t \mid\!- t + 1$ in Beobachtung treten.

In der Praxis verwenden die Versicherungsmathematiker statt der Methode der genauen Zeitdauern die ebenfalls von Ackland vorgeschlagenen Methoden der mittleren Zeitdauern (Methode DMA) oder der nächstliegenden Zeitdauern, welche Vereinfachungen der Methode der genauen Zeitdauern darstellen und sich auf dieselbe Annahme stützen wie die Methode der genauen Zeitdauern, was den Verlauf der Ausscheidewahrscheinlichkeiten während des Zeitabschnittes $t \mid\!- t + 1$ angeht, und ausserdem auf die Annahme, dass die Verteilung der Eingetretenen und der Ausgetretenen während des genannten Zeitabschnittes eine gleichmässige sei.

Wir haben die entsprechenden Formeln aufgezeigt, zu denen man gelangt, wenn man stattdessen annimmt, auf der einen Seite, dass die Ausscheideziffern der einzelnen Momente innerhalb des Zeitraumes $t \mid\!- t + 1$ konstant bleiben, und auf der anderen Seite, dass: *a*) die Differenzen zwischen den Eingetretenen und den Ausgetretenen oder Ausgeschiedenen sich gleichmässig innerhalb des Zeitabschnittes verteilen; oder *b*) während dieses Zeit-

abschnittes das Verhältnis der Differenz zwischen Eingetretenen und Ausgetretenen zu den dem Risiko ausgesetzten Personen konstant bleibe. Diese letztere Methode (Methode DMG) scheint ganz besonders geeignet für die Bestimmung der Sterblichkeit in den Erwachsenenaltern.

Ausser der Totalausscheidung, die der Gesamtheit aller in Betracht gezogenen Ausscheideursachen zuzuschreiben ist, haben wir die Teilausscheidung auf Grund einer oder mehrerer dieser Ursachen behandelt, unter gesonderter Betrachtung der reinen oder unabhängigen Wahrscheinlichkeit der Teilausscheidung und der Wahrscheinlichkeit der Teilausscheidung, die auch von den übrigen Ausscheideursachen abhängig ist. Für jeden Fall haben wir die geeigneten Formeln gegeben auf Grund der Annahme konstanter Ausscheideziffern für die einzelnen Momente. Bei dieser Annahme ergeben sich die Beziehungen zwischen reiner Wahrscheinlichkeit und abhängiger Wahrscheinlichkeit als sehr einfach.

Der Untersuchung der definitiven Ausscheideursachen haben wir eine Untersuchung der zeitweisen Ausscheideursachen folgen lassen, indem wir auch für diese geeignete Formeln vorschlugen und zur Anwendung brachten. Für die Wahrscheinlichkeit zeitweiser Ausscheidung haben wir insbesondere ausser den reinen Ausscheidewahrscheinlichkeiten die Ausscheidewahrscheinlichkeiten betrachtet, die auch von Ausscheidegründen definitiver Art abhängen.

Wenn wir uns nicht täuschen, könnten die hier vorgeschlagenen Methoden zweckmässige Anwendung finden auf dem Gebiet der verschiedenen Fächer der Statistik und insbesondere auf dem der Versicherungswissenschaften. An einem Beispiel haben wir gezeigt, wie grob im Vergleich zu den vorgenannten Methoden das Verfahren der spezifischen Koeffizienten für eheliche Fruchtbarkeit ist, das von den Volkswissenschaftlern angewandt wird, um die Fruchtbarkeit der verheirateten Frauen zu messen, und wie dasselbe zu einem von Grund aus irrigen Bilde der ehelichen Fruchtbarkeitskurve der Frau nach Altersklassen führt.

S. KULLBACK

A note on the distribution of a certain partial belonging coefficient *

In a paper in « Metron » V. ROMANOVSKY (1) considered the distribution of K. PEARSON'S (2) first partial belonging coefficient.

$$(1) \qquad \alpha_1^1 = \frac{\bar{x} - \bar{x}^1}{(S + S^1)^{\frac{1}{2}}} (m\,n)^{\frac{1}{2}}$$

where

$$\bar{x} = \frac{1}{n} \sum_{v=1}^{n} x_v \; ; \; \bar{x}^1 = \frac{1}{m} \sum_{v'=1}^{m} x_v^1 \; ; \; S = \sum_{v=1}^{n} (x_v - x_0)^2 \; ; \; S^1 = \sum_{v=1}^{m} (x_v^1 - x_0)^2 \; ,$$

and $x_1, x_2 \ldots, x_n$ and $x_1^1, x_2^1, \ldots, x_m^1$ are the values of x observed in the samples from a normal population of mean x_o and variance σ^2. ROMANOVSKY determined the moment-generating function of the distribution function of $\xi = (\bar{x} - \bar{x}^1) \left(\frac{m\,n}{m+n}\right)^{\frac{1}{2}}$ and $\eta = \frac{S + S^1}{m + n}$ but was unable to determine the distribution of ξ and η therefrom in « manageable form ». In this note we shall determine the distribution function of ξ and η by means of the theory of characteristic functions and therefrom the distribution of α_1'.

The characteristic function of the distribution function of

$$\xi = \left\{ \frac{1}{n} \sum_{v=1}^{n} (x_v - x_0) - \frac{1}{m} \sum_{v=1}^{m} (x_v^1 - x_0) \right\} \left(\frac{m\,n}{m+n}\right)^{\frac{1}{2}} \text{ and}$$

$$\eta = \frac{\sum_{v=1}^{n} (x_v - x_0)^2 + \sum_{v=1}^{m} (x_v^1 - x_0)^2}{m + n}$$

* Presented to the American Mathematical Society, June 20, 1934.

where the distribution functions of x and x^1 are respectively given by

$$(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\} \text{ and } (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(x^1-x_0)^2}{2\sigma^2}\right\} \text{ is (3)}$$

(2) $$\varphi(t_1, t_2) = I^n J^m$$

where

$$I = \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2} + \frac{it_1(x-x_0)}{n}\left(\frac{nm}{m+n}\right)^{\frac{1}{2}} + \right.$$

$$\left. + \frac{it_2(x-x_0)^2}{m+n}\right\} dx$$

and

$$J = \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(x^1-x_0)^2}{2\sigma^2} - \frac{it_1(x^1-x_0)}{m}\left(\frac{nm}{m+n}\right)^{\frac{1}{2}} + \right.$$

$$\left. + \frac{it_2(x^1-x_0)^2}{m+n}\right\} dx^1$$

and $i = \sqrt{-1}$.

Evaluating the integrals we find (1), (4)

(3) $$\varphi(t_1, t_2) = (1 - 2it_2\gamma)^{-\frac{m+n}{2}} \exp\left\{-\frac{t_1{}^2\sigma^2}{2(1-2it_2\gamma)}\right\}$$

$$\text{where } \gamma = \frac{\sigma^2}{m+n}.$$

The distribution of ξ and η is given by (3)

(4) $$P(\xi, \eta) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (1-2it_2\gamma)^{\frac{-m+n}{2}} \exp\left\{-it_1\xi - it_2\eta - \right.$$

$$\left. - \frac{t_1{}^2\sigma^2}{2(1-2it_2\gamma)}\right\} dt_1\, dt_2.$$

Since

$$\int_{-\infty}^{\infty} \exp\left\{-it_1\xi - \frac{t_1{}^2\sigma^2}{2(1-2it_2\gamma)}\right\} dt_1 =$$

$$= \left\{\frac{2\pi(1-2it_2\gamma)}{\sigma^2}\right\}^{-\frac{1}{2}} \exp\left\{-\frac{\xi^2(1-2it_2\nu)}{2\sigma^2}\right\}$$

and

$$\int_{-\infty}^{\infty} (1 - 2\,it_2\,\gamma)^{-\frac{m+n-1}{2}} \exp\left\{-it_2\left(\eta - \frac{\xi^2\gamma}{\sigma^2}\right)\right\} dt_2 =$$

$$= \frac{2\pi(2\gamma)^{-\frac{m+n-1}{2}}}{\left|\frac{m+n-1}{2}\right|}\left(\eta - \frac{\xi^2\gamma}{\sigma^2}\right)^{\frac{m+n-3}{2}} \exp\left\{-\left(\eta - \frac{\xi^2\gamma}{\sigma^2}\right)\frac{1}{2\gamma}\right\}$$

we have finally

$$(5)\quad P(\xi, \eta) = \frac{(2\pi\sigma^2)^{-\frac{1}{2}}(2\gamma)^{-\frac{m+n-1}{2}}}{\left|\frac{m+n-1}{2}\right|}\left(\eta - \frac{\xi^2\gamma}{\sigma^2}\right)^{\frac{m+n-3}{2}} e^{-\frac{\eta}{2\gamma}}$$

If we set $\theta = \frac{\xi}{\sqrt{\eta}}$ we get for the simultaneous distribution of θ and η

$$(6)\quad G(\theta, \eta) = \frac{(2\pi\sigma^2)^{-\frac{1}{2}}(2\gamma)^{-\frac{m+n-2}{2}}}{\left|\frac{m+n-1}{2}\right|}\eta^{\frac{m+n-2}{2}}\left(1 - \frac{\gamma}{\sigma^2}\theta^2\right)^{\frac{m+n-3}{2}} e^{-\frac{\eta}{2\gamma}}$$

The distribution of θ is given by $D(\theta) = \int_0^{\infty} G(\theta, \eta)\, d\eta$ so that

$$(7)\qquad D(\theta) = \frac{\left|\frac{m+n}{2}\right|(2\gamma)^{\frac{1}{2}}}{(2\pi\sigma^2)^{\frac{1}{2}}\left|\frac{m+n-1}{2}\right|}\left(1 - \frac{\gamma\theta^2}{\sigma^2}\right)^{\frac{m+n-3}{2}}$$

or

$$(8)\quad D(\alpha_1') = \frac{\left|\frac{m+n}{2}\right|}{(m+n)^{\frac{1}{2}}\sqrt{\pi}\left|\frac{m+n-1}{2}\right|}\left(1 - \frac{\alpha_1'^2}{m+n}\right)^{\frac{m+n-3}{2}},\quad \alpha_1'^2 \leqq m+n\,.$$

Since

$$\left(\alpha_1'^{2p+1}\right) = \frac{\left|\frac{m+n}{2}\right|}{(m+n)^{\frac{1}{2}}\left|\frac{m+n-1}{2}\right|\sqrt{\pi}} \int_{-\sqrt{m+n}}^{\sqrt{m+n}} \alpha_1'^{2p+1}\left(1 - \frac{\alpha_1'^2}{m+n}\right)^{\frac{m+n-3}{2}} d\alpha_1'$$

we have that

$$E\left(\alpha_1'^{2p+1}\right) = 0 .$$

Also

$$E\left(\alpha_1'^{2p}\right) = \frac{\left|\frac{m+n}{2}\right|}{(m+n)^{\frac{1}{2}}\left|\frac{m+n-1}{2}\right|\sqrt{\pi}} \int_{-\sqrt{m+n}}^{\sqrt{m+n}} \alpha_1'^{2p}\left(1-\frac{\alpha_1'^2}{m+n}\right)^{\frac{m+n-3}{2}} d\alpha_1'$$

$$= \frac{(m+n)^p\left|\frac{m+n}{2}\right|}{\left|\frac{m+n-1}{2}\right|\sqrt{\pi}} \int_0^1 y^{\frac{2p-1}{2}}(1-y)^{\frac{m+n-3}{2}} d$$

$$= \frac{(m+n)^p\left|\frac{m+n}{2}\right|}{\sqrt{\pi}\left|\frac{m+n+2p}{2}\right|} \cdot \frac{2p-1}{2}\cdot\frac{2p-3}{2}\cdots\frac{1}{2}\left|\frac{1}{2}\right| =$$

$$= \frac{\left|\frac{m+n}{2}\right|(m+n)^p(2p)!}{\left|\frac{m+n+2p}{2}\right| 2^{2p}\cdot p!}$$

$$= \frac{(2p)!\left|s/2\right|}{2^p p!\left|\frac{s+2p}{2}\right|}\left(\frac{s}{2}\right)^p$$

where $S = m + n$.

If we write (8) in the form

$$(9) \qquad D(\alpha_1') = \frac{\left|s/2\right|}{s^{1/2}\left|\frac{s-1}{2}\right|\sqrt{\pi}}\left(1-\frac{\alpha_1'^2}{s}\right)^{\frac{s-3}{2}};$$

$$\text{since } \left|s/2\right| \Big/ s^{\frac{1}{2}}\left|\frac{s-1}{2}\right| = 2^{-\frac{1}{2}}\left(1-\frac{3}{4s}+\ldots\right)$$

we have that $D(\alpha_1')$ tends to $(2\pi)^{-\frac{1}{2}}\ ex\ p\left\{-\frac{\alpha_1'^2}{2}\right\}$ as $s \to \infty$.

These results, as to the moments and limit agree with those obtained by ROMANOVSKY in a different manner.

In connection with the preceding it is of interest that the simultaneous distribution of

$$u_1 = \sum_{v=1}^{n} (x_v - x_o) \text{ and } u_2 = \sum_{v=1}^{n} (x_v - x_o)^2$$

where the $x's$ are distributed according to

$$b(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(x-x_o)^2}{2\sigma^2}\right\}$$

is found to be by a method similar to that used above (4)

$$P(u_1, u_2) = \frac{(u_2 - u_1^2/n)^{\frac{n-3}{2}} e^{-\frac{u_2}{2\sigma^2}}}{\sqrt{2\pi n}\, 2^{\frac{n-1}{2}} \sigma^n \left|\frac{n-1}{2}\right|} \tag{10}$$

From this we easily find the distribution of $\theta = \frac{u_1}{\sqrt{u_2}}$ to be given by

$$D(\theta) = \frac{\left|\frac{n/2}{}\right|}{n^{1/2}\sqrt{\pi}\left|\frac{n-1}{2}\right|}\left(1 - \frac{\theta^2}{n}\right)^{\frac{n-3}{2}} \tag{11}$$

But $\theta^2 = u_1^2/u_2 = \frac{n^2(\bar{x}-x_0)^2}{n s^2 + n(\bar{x}-x_0)^2} = \frac{n z^2}{1+z^2}$ where $z^2 = \frac{(\bar{x}-x_0)^2}{s^2}$

and s^2 is the sample variance. Applying this transformation to (11) we find for the distribution of z, the well known « Student's » distribution

$$P(z) = \frac{\left|\frac{n/2}{}\right|}{\sqrt{\pi}\left|\frac{n-1}{2}\right|}(1+z^2)^{-\frac{n}{2}} \tag{12}$$

It is clear that (9) may be changed to this form by the transformation $\frac{\alpha_1'^2}{s} = \frac{t^2}{1+t^2}$.

REFERENCES

(1) V. Romanovsky. *On the criteria that two given samples belong to the same population.* « Metron » Vol. 7 (1928), N. 3, pp. 3-46.

(2) K. Pearson. *On the coefficient of racial likeness.* « Biometrika », Vol. 18 (1926), pp. 105-117.

(3) S. Kullback. *An application of characteristic functions to the distribution problem of statistics.* « Annals of Mathematical Statistics » Vol. 5 (1934) pp. 263-307..

(4) Wishart and Bartlett. *The distribution of second order moment statistics in a normal system.* « Proc. Camb. Phil. Soc. » Vol. 28 (1932), p. 455, Equation 6.

A. LINDER

" Wahrscheinlichkeitsansteckung ,, und Differenzengleichungen

Wie GULDBERG in mehreren Arbeiten am Beispiel der Bernoullischen, Poissonschen, Pascalschen und *hypergeometrischen* Verteilungen überzeugend dargetan hat (1), (2), (3) (*), lassen sich wichtige, in der Theorie der ein- und der mehrdimensionalen statistischen Verteilungen auftretende Probleme mit Hülfe der Differenzengleichungen in einfacher Weise behandeln. RAGNAR FRISCH hat dann die Guldbergsche Theorie in interessanter Weise verallgemeinert (4).

In dieser Arbeit soll gezeigt werden, dass die Methode von Guldberg auf den Fall der *Wahrscheinlichkeitsansteckung* ausgedehnt werden kann. Das Schema der Wahrscheinlichkeitsansteckung stammt von POLYA (5) und ist von EGGENBERGER (6) gründlich behandelt worden. Die von GULDBERG herangezogenen Bernoullischen, Poissonschen und hypergeometrischen Verteilungen erscheinen als Sonderfälle der Verteilung, die sich aus dem allgemeinen Schema der Wahrscheinlichkeitsansteckung ergibt.

1. — DAS SCHEMA DER WAHRSCHEINLICHKEITSANSTECKUNG.

In einer Urne befinden sich k weisse und h schwarze Kugeln. Wir ziehen m mal eine Kugel und legen jedesmal $1 + d$ Kugeln

(*) Die Literatur ist am Schlusse der Arbeit zusammengestellt.

Mehrere Monate nach der Absendung der vorliegenden Arbeit an das « Metron » erhielt ich Kenntnis von einer Arbeit von OTA FISCHER (*Une remarque sur l'article de M. A. Guldberg : « On discontinuous frequency-functions and statistical series »*. « Aktuarské Vedy », IV, 1933-934, S. 169-174), in welcher einige der von mir gefundenen Formeln ebenfalls gegeben werden.

von der gezogenen Farbe in die Urne zurück. Die Wahrscheinlichkeit, $f(x)$, in m Zügen x weisse Kugeln zu ziehen, ist gegeben durch

$$(1)\quad f(x) = \binom{m}{x} \frac{k(k+d)(k+2d)\dots(k+(x-1)d)\,h(h+d)\dots(h+(m-x-1)d)}{(k+h)(k+h+d)(k+h+2d)\dots(k+h+(m-1)d)}$$

Setzen wir in dieser Formel $d = 0$, so erhalten wir die Wahrscheinlichkeit dafür, in m Zügen x mal eine weisse Kugel zu ziehen, wenn jedesmal die gezogene Kugel in die Urne zurückgelegt wird: *Bernoullisches Urnenschema.* Aus (1) wird, wenn noch

$$\frac{k}{k+h} = p \qquad \text{und} \qquad \frac{h}{k+h} = 1-p$$

gesetzt wird

$$f(x) = \binom{m}{x} p^x (1-p)^{m-x}$$

Setzen wir $d = -1$, so erhalten wir das Urnenschema für die *hypergeometrische Verteilung.* $f(x)$ stellt uns die Wahrscheinlichkeit dar, in m Zügen x weisse Kugeln zu ziehen, wenn bei jedem Zug die Kugel *nicht in die Urne zurückgelegt* wird. Für $d = -1$ wird aus (1):

$$f(x) = \binom{m}{x} \frac{k(k-1)(k-2)\dots(k-(x-1))\,h(h-1)(h-2)\dots(h-(m-x-1))}{(k+h)(k+h-1)(k+h-2)\dots(k+h-(m-1))}$$

$$= \frac{m!}{x!\,(m-x)!} \cdot \frac{k!}{(k-x)!} \cdot \frac{h!}{(h-(m-x))!} \cdot \frac{(k+h-m)!}{(k+h)!}$$

$$f(x) = \frac{\binom{k}{x}\binom{h}{m-x}}{\binom{k+h}{m}}$$

Bekanntlich erhält man aus der Bernoullischen Verteilung die Poissonsche durch einen Grenzübergang, den man als den Uebergang zu den « seltenen Fällen » bezeichnen kann. Wie Eggenberger gezeigt hat, kann man aus (1) durch einen ebensolchen Uebergang zu einer Verteilung gelangen, die der Poissonschen entspricht. Man hat nur

$$\frac{mk}{h+k} = \lambda \qquad \text{fest anzunehmen}$$

und m gegen ∞ streben zu lassen. $\frac{k}{h+k}$ strebt gegen Null. Damit

ist gefordert, dass es sich um « *seltene Ereignisse* » handelt. Fordert man dazu « *schwache Chancenvermehrung* », so muss

$$\frac{m\,d}{h+k} = \delta \qquad \text{fest}$$

und $\frac{d}{h+k}$ gegen Null strebend angenommen werden. Unter diesen Bedingungen erhält man aus (1) (siehe EGGENBERGER, Formel (20)) :

$$(1') \qquad f(x) = \frac{\lambda\,(\lambda+\delta)\,(\lambda+2\,\delta)\ldots(\lambda+(x-1)\,\delta)}{x!\quad (1+\delta)^{\frac{\lambda}{\delta}+x}}$$

Wähot man λ wie oben und $\delta = 0$, so erhält man aus (1)

$$f(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

die bekannte Formel für die Poissonsche Verteilung.

2. — Die Differenzengleichung.

Aus Gleichung (1) erhält man leicht die folgende Differenzengleichung der Verteilung für die Wahrscheinlichkeitsansteckung.

$$(2) \qquad f(x+1) = \frac{(m-x)\,(k+x\,d)}{(x+1)\,(h+(m-x-1)\,d)}\,f(x)$$

Diese Differenzengleichung findet sich schon bei EGGENBERGER, doch ohne die Folgerungen, die wir daraus ziehen werden.

Setzen wir in (2) $d = 0$, so erhalten wir die Differenzengleichung für die *Bernoullische Verteilung*.

$$f(x+1) = \frac{(m-x)\,k}{(x+1)\,.\,h}\,.\,f(x)$$

Schreiben wir für

$$\frac{k}{k+h} = p \qquad \text{und} \qquad \frac{h}{h+k} = 1-p$$

so wird daraus

$$f(x+1) = \frac{p}{1-p}\cdot\frac{m-x}{x+1}\,f(x)$$

die Differenzengleichung, die GULDBERG (1) auf Seite 172 anführt.

Nehmen wir in (2) $d = -1$ an, so ergibt sich die Differenzengleichung für die *hypergeometrische Verteilung* in der Form

$$f(x+1) = \frac{(m-x)\,(k-x)}{(x+1)\,(h-m+x+1)}\,f(x)$$

Es ist dies die Formel, die sich bei GULDBERG (1) auf Seite 179 findet.

Aus (2) erhalten wir auch die Differenzengleichung für die Verteilung im Falle der *seltenen Ereignisse mit schwacher Chancenvermehrung*. Da wir in diesem Falle

$$\frac{m \,.\, k}{h + k} = \lambda = \text{fest, wenn } m \to \infty,$$

$$\frac{m \,.\, d}{h + k} = \delta = \text{fest, wenn } m \to \infty,$$

sowie $\frac{k}{h+k}$ und $\frac{d}{h+k}$ gegen Null strebend annehmen müssen, erhalten wir aus (2), wenn noch

$$\frac{h}{h+k} = 1 - \frac{k}{h+k} \text{ gesetzt wird,}$$

$$f(x+1) = \frac{\left(1 - \frac{x}{m}\right)\left(\frac{m\,k}{h+k} + x\,\frac{m\,d}{h+k}\right)}{(x+1)\left(1 - \frac{k}{h+k} + \left(1 - \frac{x+1}{m}\right)\delta\right)}$$

Für $m \to \infty$ wird daraus

$$f(x+1) = \frac{\lambda + x \,.\, \delta}{(x+1)(1+\delta)} f(x) \tag{2'}$$

als Differenzengleichung für den Fall der seltenen Ereignisse mit schwacher Chancenvermehrung. Auch diese Gleichung findet sich bei EGGENBERGER[(6), Formel (23)]. Aus ihr finden wir für $\delta = 0$:

$$f(x+1) = \frac{\lambda}{x+1} f(x)$$

die Differenzengleichung für die *Poissonsche Verteilung* (siehe GULDBERG, (1), S. 173).

3. — REKURSIONSFORMELN FÜR DIE MOMENTE.

Das r — te Moment um den Ursprung ist definiert durch

$$\sigma_r = \sum_{x=0}^{k} x^r f(x)$$

Mit Hülfe der Differenzengleichung (2) lässt sich eine Rekursionsformel für σ_n ableiten. Wir schreiben die Differenzengleichung (2) in der Form

$$(h + m\,d)\,(x+1)\,f(x+1) - d\,(x+1)^2 f(x+1) =$$
$$= m\,k\,f(x) + (m\,d - k)\,x\,f(x) - d\,x^2 f(x)$$

Die linke Seite dieser Gleichung wird mit $(x+1)^n$, die rechte mit

$$x^n + \binom{n}{1} x^{n-1} + \binom{n}{2} x^{n-2} + \ldots + x + 1$$

multipliziert.

$$(h + m\,d)\,(x+1)^{n+1} f(x+1) - d\,(x+1)^{n+2} f(x+1) =$$
$$= m\,k\,x^n f(x) + m\,k \binom{n}{1} x^{n-1} f(x) + \ldots + m\,k\,f(x)$$
$$+ (m\,d - k)\,x^{n+1} f(x) + (m\,d - k)\binom{n}{1} x^n + \ldots + (m\,d - k)\,x\,.\,f(x)$$
$$- d\,x^{n+2} f(x) - d\binom{n}{1} x^{n+1} f(x) - \ldots - d\,x^2 f(x)$$

Durch Summation von o bis k erhält man

$$(h + m\,d)\,\sigma_{n+1} - d\,\sigma_{n+2} =$$
$$= m\,k\,\sigma_n + m\,k\binom{n}{1}\sigma_{n-1} + m\,k\binom{n}{2}\sigma_{n-2} + \ldots + m\,k$$
$$+ (m\,d - k)\,\sigma_{n+1} + (m\,d - k)\binom{n}{1}\sigma_n + (m\,d - k)\binom{n}{2}\sigma_{n-1} + \ldots + (m\,d - k)\,\sigma_1$$
$$- d\,\sigma_{n+2} - d\binom{n}{1}\sigma_{n+1} - d\binom{n}{2}\sigma_n - \ldots - d\,\sigma_2$$

Die Rekursionsformel für die Momente lautet endgültig

$$(h + k + \binom{n}{1} d)\,\sigma_{n+1} = m\,k\,\sigma_n + m\,k\binom{n}{1}\sigma_{n-1} + m\,k\binom{n}{2}\sigma_{n-2} + \ldots + m\,k$$
$$+ (m\,d - k)\binom{n}{1}\sigma_n + (m\,d - k)\binom{n}{2}\sigma_{n-1} + (m\,d - k)\binom{n}{3}\sigma_{n-2} + \ldots + (m\,d - k)\,\sigma_1 \quad (3)$$
$$- d\binom{n}{2}\sigma_n - d\binom{n}{3}\sigma_{n-1} - d\binom{n}{4}\sigma_{n-2} - \ldots - d\,\sigma_2$$

Für $n = 0$ erhält man

$$\sigma_1 = m\,\frac{k}{h+k}$$

Für $n = 1$:

$$(h + k + d)\,\sigma_2 = m\,k\,\sigma_1 + m\,k + (m\,d - k)\,\sigma_1$$

$$\sigma_2 = \frac{m\,k\,(m\,(k + d) + h)}{(h + k)\,(h + k + d)}$$

Setzen wir in (3) $d = 0$, so erhalten wir die von GULDBERG angegebene Rekursionsformel für die binomische Verteilung, für $d = -1$ diejenige für die hypergeometrische Verteilung.

Unter den bekannten Voraussetzungen für *seltene Ereignisse und schwache Chancenvermehrung* wird aus (3)

$$\sigma_{n+1} = \lambda\,\sigma_n + \binom{n}{1}\lambda\,\sigma_{n-1} + \binom{n}{2}\lambda\,\sigma_{n-2} + \ldots + \lambda$$

(3′)

$$+ \binom{n}{1}\delta\,\sigma_n + \binom{n}{2}\delta\,\sigma_{n-1} + \binom{n}{3}\delta\,\sigma_{n-2} + \ldots + \delta\,\sigma_1 .$$

Für $\delta = 0$ erhält man wiederum die entsprechende Formel für die Poissonsche Verteilung.

Wir gehen über zur Berechnung der Rekursionsformel für die *unvollständigen Momente*. Dieselben sind durch folgende Beziehung definiert:

$${}_t\sigma_r = \sum_{x=t}^{k} x^r\,f\,(x)$$

Aus der selbstverständlichen Gleichung

$$\sum_{x=t}^{k} x^{n+1}\,f\,(x) = \sum_{x=t}^{k} (x + 1)^{n+1}\,f\,(x + 1) + t^{n+1}\,f\,(t)$$

folgt

$$\sum_{x=}^{k} (x + 1)^{n+1}\,f\,(x + 1) = {}_t\sigma_{n+1} - t^{n+1}\,f\,(t)$$

Summieren wir in der dritten Gleichung des vorigen Abschnitts über x von t bis k, so erhalten wir bei Beachtung der letzten Gleichung

$$(h + m\,d)\,({}_t\sigma_{n+1} - t^{n+1}\,f\,(t)) - d\,({}_t\sigma_{n+2} - t^{n+2}\,f\,(t)) =$$

$$= m\,k\,{}_t\sigma_n + \binom{n}{1} m\,k\,{}_t\sigma_{n-1} + \ldots \qquad + m\,k$$

$$+ (m\,d - k)\,{}_t\sigma_{n+1} + \binom{n}{1}(m\,d - k)\,{}_t\sigma_n + \ldots \quad + (m\,d - k)\,{}_t\sigma_1$$

$$- d\,{}_t\sigma_{n+2} - \binom{n}{1} d\,{}_t\sigma_{n+1} - \ldots \qquad - d\,{}_t\sigma_2$$

und daraus

$$(h + k + \binom{n}{1} d)\, {}_t\sigma_{n+1} = (h + m\,d)\, t^{n+1} f(t) - d\, t^{n+2} f(t)$$
$$+ m\,k\, {}_t\sigma_n + \binom{n}{1} m\,k\, {}_t\sigma_{n-1} + \ldots \quad + m\,k$$
(4)
$$+ \binom{n}{1} (m\,d - k)\, {}_t\sigma_n + \binom{n}{2} (m\,d - k)\, {}_t\sigma_{n-1} + \ldots + (m\,d - k)\, {}_t\sigma_1$$
$$- \binom{n}{2} d\, {}_t\sigma_n - \binom{n}{3} d\, {}_t\sigma_{n-1} - \ldots \quad - d\, {}_t\sigma_2$$

Für $d = 0$ geht (4) in die von GULDBERG angegebene Gleichung für die Bernoullische Verteilung über. Ebenso geht sie für $d = -1$ in die Rekursionsformel für die hypergeometrische Verteilung über, die bei GULDBERG für $n = o$ angegeben wird.

Im Falle der *seltenen Ereignisse mit schwacher Chancenvermehrung* erhalten wir durch Ausfürung des schon erwähnten Grenzüberganges

$${}_t\sigma_{n+1} = (1 + \delta)\, t^{n+1} f(t)$$
$$+ \lambda\, {}_t\sigma_n + \binom{n}{1} \lambda\, {}_t\sigma_{n-1} + \ldots \quad + \lambda$$
(4')
$$+ \binom{n}{1} \delta\, {}_t\sigma_n + \binom{n}{2} \delta\, {}_t\sigma_{n-1} + \ldots \quad + \delta\, {}_t\sigma_1$$

Aus dieser Formel (4') erhalten wir für $\delta = 0$ die von GULDBERG angeführte Rekursionsformel für die unvollständigen Momente der *Poissonschen* Verteilung.

4. — DIE LOKALEN UND DIE TOTALEN KRITERIEN FÜR DIE WAHRSCHEINLICHKEITSANSTECKUNG.

Die *lokalen* Kriterien geben an, wie sich die theoretische Verteilung der empirischen in jedem Punkt x anschliesst. Zur Ableitung des lokalen Kriteriums für den allgemeinen Fall der Wahrscheinlichkeitsansteckung gehen wir von (3) aus, indem wir nacheinander $n = 0, 1, 2$ setzen.

$$\cdot \frac{m\,k}{h + k} = \sigma_1$$
$$\cdot \frac{m\,k}{h + k} (\sigma_1 + 1) + \frac{m\,d - k}{h + k} \sigma_1 - \frac{d}{h + k} \sigma_2 = \sigma_2$$
$$\cdot \frac{m\,k}{h + k} (\sigma_2 + 2\,\sigma_1 + 1) + \frac{m\,d - k}{h + k} (2\,\sigma_2 + \sigma_1) - \frac{d}{h + k} (2\,\sigma_3 + \sigma_2) = \sigma_3$$

Aus diesen drei Gleichungen bestimmen wir

$$\frac{m\,k}{h+k}, \frac{m\,d-k}{h+k} \text{ und } \frac{d}{h+k}.$$

$$\frac{m\,k}{h+k} = \sigma_1$$

$$\frac{k-m\,d}{h+k} = \frac{\sigma_2(\sigma_2-\sigma_3)+\sigma_1[\sigma_2^2-2\,\sigma_3(1+\sigma_1)]}{2(\sigma_2^2-\sigma_1\sigma_3)}$$

$$\frac{d}{h+k} = \frac{\sigma_1(\sigma_3+\sigma_2+\sigma_1\sigma_2)-2\,\sigma_2^2}{2(\sigma_2^2-\sigma_1\sigma_3)}$$

Die Differenzengleichung (2) schreiben wir in der Form

$$\frac{f(x+1)}{f(x)}(x+1)\frac{h+m\,d}{h+k} - \frac{f(x+1)}{f(x)}(x+1)^2\frac{d}{h+k} + \\ + \frac{k-m\,d}{h+k}x + \frac{d}{h+k}x^2 = \frac{k\,m}{h+k}$$

Es ist noch $\frac{h+m\,d}{h+k}$ zu bestimmen. Nun ist aber

$$\frac{m\,d+h}{h+k} = \frac{m\,d-k}{h+k} + \frac{h+k}{h+k} = \frac{m\,d-k}{h+k} + 1$$

Setzen wir darin den oben für $\frac{m\,d-k}{h+k}$ erhaltenen Wert ein, so wird

$$\frac{m\,d+h}{h+k} = \frac{\sigma_2(\sigma_2+\sigma_3)+\sigma_1(2\,\sigma_1\sigma_3-\sigma_2^2)}{2(\sigma_2^2-\sigma_1\sigma_3)}$$

Wir können nun die Konstanten der Differenzengleichung als Funktionen der Momente schreiben und erhalten

$$\frac{f(x+1)}{f(x)}(x+1)[\sigma_2(\sigma_2+\sigma_3)+\sigma_1(2\,\sigma_1\sigma_3-\sigma_2^2)] - \\ - \frac{f(x+1)}{f(x)}(x+1)^2[\sigma_1(\sigma_2+\sigma_3+\sigma_1\sigma_2)-2\,\sigma_2^2] \\ + [\sigma_2(\sigma_2-\sigma_3)+\sigma_1\{\sigma_2^2-2\,\sigma_3(1+\sigma_1)\}]x + \\ + [\sigma_1(\sigma_3+\sigma_2+\sigma_1\sigma_2)-2\,\sigma_2^2]x^2 \\ = 2\,\sigma_1(\sigma_2^2-\sigma_1\sigma_3)$$

Wir ersetzen die Momente um den Ursprung durch die Semiinvarianten von Thiele nach den Formeln (siehe GULDBERG, (2), S. 234)

$$\sigma_1 = \mu_1$$

$$\sigma_2 = \mu_2 + \mu_1^2$$

$$\sigma_3 = \mu_3 + \mu_1 (3\,\mu_2 + \mu_1^2)$$

Dann erhalten wir

$$\frac{f(x+1)}{f(x)}(x+1)\left\{[\mu_3 + \mu_1(3\,\mu_2 + \mu_1^2)][\mu_2 + 3\,\mu_1^2] + (1-\mu_1)(\mu_2 + \mu_1^2)^2\right\}$$

$$(5)\quad -\frac{f(x+1)}{f(x)}(x+1)^2[\mu_1(\mu_3 + \mu_2 + \mu_1^2) - 2\,\mu_2^2] + x^2[\mu_1(\mu_3 + \mu_2 + \mu_1^2) - 2\,\mu_2^2]$$

$$+ x\left\{(1+\mu_1)(\mu_2 + \mu_1^2)^2 - [\mu_3 + \mu_1(3\,\mu_2 + \mu_1^2)][\mu_2 + 3\,\mu_1^2 + 2\,\mu_1]\right\}$$

$$= 2\,\mu_1[\mu_2^2 - \mu_1(\mu_1\,\mu_2 + \mu_3)]$$

Bezeichnen wir die linke Seite der Gleichung (5) mit $\Psi(x)$, so ist

$$(6)\qquad \alpha(x) = \frac{\Psi(x)}{2\,\mu_1[\mu_2^2 - \mu_1(\mu_1\,\mu_2 + \mu_3)]} = 1$$

das *lokale Kriterium* für die Verteilung gemäss dem allgemeinen Schema der Wahrscheinlichkeitsansteckung.

Das lokale Kriterium vereinfacht sich ganz erheblich, wenn wir zum Schema der *seltenen Ereignisse mit schwacher Chancenvermehrung* übergehen. Wir stützen uns in diesem Falle auf die Gleichung (3′) und erhalten

$$\sigma_1 = \lambda$$

$$\sigma_2 = \lambda\,\sigma_1 + \lambda + \delta\,\sigma_1$$

oder also

$$\lambda = \sigma_1$$

$$\delta = \frac{\sigma_2 - \sigma_1^2 - \sigma_1}{\sigma_1}$$

Durch Einsetzen der Werte von λ und δ in die Differenzengleichung (2′), die wir wie folgt schreiben

$$\frac{f(x+1)}{f(x)}(x+1)(1+\delta) - x\,\delta = \lambda$$

erhält man

$$\frac{f(x+1)}{f(x)}(x+1)(\sigma_2-\sigma_1^2)-x(\sigma_2-\sigma_1-\sigma_1^2)=\sigma_1^2$$

In dieser Formel können wieder anstelle der Momente um den Ursprung die Semiinvarianten eingeführt werden, was zu

$$\frac{f(x+1)}{f(x)}(x+1)+x\frac{\mu_1-\mu_2}{\mu_2}=\frac{\mu_1^2}{\mu_2} \tag{5'}$$

führt. Diese Gleichung entspricht genau der von GULDBERG für die Bernoullische, Poissonsche und Pascalsche Verteilung gefundenen, wobei

$$\mu_1 \gtreqless \mu_2 .$$

In (5') ist

$$\mu_1 = \lambda$$
$$\mu_2 = \lambda(1+\delta)$$

Es ist also

$$\mu_1 \gtreqless \mu_2 , \quad \text{jenachdem} \quad \delta \lesseqgtr 0 .$$

Somit entsprechen sich die Differenzengleichungen und damit die lokalen Kriterien für

Wahrscheinlichkeitsansteckung, seltene Ereignisse, schwache Chancenvermehrung	Unabhängige Wahrscheinlichkeiten
$\delta < 0$	Bernoullischr Verteilung
$\delta = 0$ identisch mit . .	Poissonscher »
$\delta > 0$	Pascalsche »

In allen diesen Fällen ist also der *Approximationskoeffizient*

$$\alpha(x) = \frac{\Psi(x)}{\frac{\mu_1^2}{\mu_2}} \tag{6'}$$

$$\text{mit} \quad \Psi(x) = \frac{f(x+1)}{f(x)}(x+1)+x\frac{\mu_2-\mu_1}{\mu_2}$$

Für die *totalen Kriterien* ergeben sich im allgemeinen Falle der Wahrscheinlichkeitsansteckung ziemlich umfangreiche Ausdrücke, auf die wir hier wegen ihrer Schwerfälligkeit für praktische Bedürfnisse nicht weiter eingehen.

Im Falle seltener Ereignisse und schwacher Chancenvermehrung ergeben sich auch die totalen Kriterien in sehr einfacher Weise. Die Semiinvarianten lassen sich mit Hülfe der Gleichung (3′) und der Beziehung

$$\sigma_{r+1} = \mu_1 \sigma_r + \binom{r}{1} \mu_2 \sigma_{r-1} + \binom{r}{2} \mu_3 \sigma_{r-2} + \ldots + \mu_{r+1}$$

(siehe GULDBERG, (2), S. 234 unterste Formel) mühelos finden. Man erhält für die vier ersten Semiinvarianten

$$\mu_1 = \lambda$$
$$\mu_2 = \lambda (1 + \delta)$$
$$\mu_3 = \lambda (1 + \delta) (1 + 2\delta)$$
$$\mu_4 = \lambda (1 + \delta) (1 + 6\delta (1 + \delta))$$

Aus den ersten drei Gleichungen ergibt sich

$$(7'a) \qquad \frac{\mu_1 \mu_3}{2\mu_2^2 - \mu_1 \mu_2} = 1$$

Aus den zwei ersten und der vierten

$$(7'b) \qquad \frac{\mu_1^2 \mu_4}{\mu_1^2 \mu_2 - 6\mu_1 \mu_2^2 + 6\mu_2^3} = 1$$

Es sind dies genau dieselben Gleichungen, wie sie GULDBERG für die Bernoullische Verteilung hergeleitet hat ((1), S. 173). Wie man sich leicht überzeugt, gelten dieselben Gleichungen auch für die Poissonsche und die Pascalsche Verteilung.

Aus den beiden ersten *totalen Kriterien* ergibt sich somit noch nicht, ob eine gegebene Zahlentafel einer der einfachen (Bernoullischen, Poissonschen oder Pascalschen) Verteilungen entspricht, oder ob es sich um eine Verteilung gemäss dem Schema der seltenen Ereignisse mit schwacher Chancenvermehrung handelt. Da auch die *lokalen Kriterien* sich nicht voneinander unterscheiden, muss der Bearbeiter einer Statistik durch sachliche, in der Natur der beschriebenen Erscheinung liegende Gründe zu beurteilen suchen, welche der theoretisch möglichen Verteilungen in Frage kommt.

Ist also, wie in dem von EGGENBERGER behandelten Beispiel, auf das wir im nächsten Abschnitt näher eintreten, $\mu_1 < \mu_2$, so kann theoretisch bei Erfüllung der lokalen und der totalen Kriterien ebensogut eine Pascalsche Verteilung als eine Verteilung mit positiver Wahrscheinlichkeitsansteckung ($\delta > 0$) zugrunde liegen.

Da es sich in diesem Beispiel um Todesfälle an einer ansteckenden Krankheit handelt (Scharlach), wird man sich wohl ohne weiteres für die zweite Lösung entschliessen (Siehe Beispiel *d*).

5. — Beispiele.

a) *Häufigkeit der Todesfälle infolge Stich oder Biss giftiger Tiere* (*ausgenommen Lyssa*) *in der Schweiz in den Jahren* 1877 *bis* 1933. — Die Zahlen des ersten Beispiels entnehme ich einer Zusammenstellung des *Eidgenössischen Statistischen Amtes*, die mir in liebenswürdiger Weise Herr Professor Dr. W. Friedli in Bern zur Verfügung stellte. Die Todesfälle infolge Stich oder Biss giftiger Tiere sind in der Schweiz selten. An dieser Todesursache sind in den letzten 57 Jahren jährlich durchschnittlich nur $5\frac{1}{2}$ Personen gestorben.

Wir stellen zunächst in einer Tabelle die empirischen und die theoretischen Zahlen der Verteilung zusammen, denen wir noch die Approximationskoefflzienten beifügen.

Zahl der Todesfälle in 1 Jahr x	Anzahl der Jahre mit x Todesfällen $H(x)$	Approximations-koeffizient $\alpha(x)$	Nach Formel (5′) berechnete Häufigkeit $57\,f(x)$
0	1	1,1	1,3
1	3	0,8	3,5
2	4	2,1	5,7
3	9	0,9	7,1
4	9	0,9	7,5
5	8	1,0	7,1
6	7	0,4	6,2
7	4	0,9	5,0
8	3	1,8	3,9
9	3	— 0,4	2,9
10	1	— 1,8	2,1
11	—	∞	1,5
12	2	0,2	1,0
13	1	— 2,4	0,7
14	—	∞	0,4
15	2	— 2.8	0,3

Die Uebereinstimmung zwischen der empirischen und der theoretischen Verteilung ist befriedigend, was auch im Pendeln der Approximationskoefflzienten um den Wert 1 zum Ausdruck kommt.

Für die vier ersten Semiinvarianten erhält man folgende Werte :

$$\mu_1 = 5{,}47$$
$$\mu_2 = 10{,}99$$
$$\mu_3 = 39{,}28$$
$$\mu_4 = 325{,}24$$

Daraus ergibt sich für die beiden ersten der totalen Kriterien :

$$\frac{\mu_1 \mu_3}{2\mu_2^2 - \mu_1 \mu_2} = 1{,}2$$

$$\frac{\mu_1^2 \mu_4}{\mu_1^2 \mu_2 - 6\mu_1 \mu_2^2 + 6\mu_2^3} = 2{,}2$$

Die für die Verteilung charakteristischen Konstanten bestimmt man zu

$$\lambda = \mu_1 = 5{,}47$$

$$\delta = \frac{\mu_2 - \mu_1}{\mu_1} = 1{,}01$$

b) *Häufigkeit der Selbstmorde weiblicher Personen in der Schweiz, in den Jahren* 1901-1920. — In einer Veröffentlichung des Eidgenössischen Statistischen Amtes (7) sind die Selbstmordtodesfälle von Personen weiblichen Geschlechts in den Jahren 1901-1920 nach Monaten angegeben. Auf einen Monat entfallen durchschnittlich rund 14 Fälle. Die Verteilung auf die einzelnen Monate lässt sich dusch eine Verteilung gemäss dem Schema der Wahrscheinlichkeitsansteckung gut wiedergeben.

Zahl der Todesfälle in 1 Monat x	Anzahl der Monate mit x Todesfällen $H(x)$	Approximations-koeffizient $\alpha(x)$	Nach Formel (5′) berechnete Häufigkeit $240\, f(x)$
4	1	1,5	1,3
5	3	1,8	2,6
6	9	0,3	4,7
7	6	1,6	7,5
8	13	0,8	10,7
9	14	0,8	14,0
10	14	1,5	16,9
11	22	0,5	19,2
12	15	1,4	20,4
13	19	1,2	20,6
14	21	1,4	19,8
15	25	0,4	18,3
16	14	1,1	16,2
17	13	1,2	13,8
18	12	0,9	11,4
19	9	1,7	9,1
20	10	0,6	7,0
21	6	2,2	5,3
22	7	— 0,1	3,9
23	2	1,8	2,8
24	2	1,8	2,0
25	2	— 0,9	1,4
26	—	2,0	0,9
27	—	∞	0,6
28	1	— 1,0	0,4

Die ersten vier Semiinvarianten, die Konstanten der Verteilung und die totalen Kriterien nehmen folgende Werte an :

$\mu_1 = 13{,}78$

$\mu_2 = 20{,}99$

$\mu_3 = 26{,}40$

$\mu_4 = -163{,}83$

$\lambda = 13{,}78$

$\delta = 0{,}52$

$$\frac{\mu_1 \mu_3}{2\mu_2^2 - \mu_1 \mu_2} = 0{,}61$$

$$\frac{\mu_1^2 \mu_4}{\mu_1^2 \mu_2 - 6\mu_2^2 \mu_1 + 6\mu_2^3} = 1{,}04$$

c) *Häufigkeit der Todesfälle infolge Erfrierens von Personen weiblichen Geschlechts in der Schweiz während der Jahre* 1877-1933. — In der Schweiz sterben jährlich im Durchschnitt 4 Personen weiblichen Geschlechts an Erfrieren. Auch diese Verteilung habe ich einer Zusammenstellung des *Eidgenössischen Statistischen Amtes* entnommen, die ich Herrn Prof. FRIEDLI verdanke. Sie lässt sich ebenfalls recht gut durch das Schema der Wahrscheinlichkeitsansteckung wiedergeben.

Zahl der Todesfälle in 1 Jahr x	Anzahl der Jahre mit x Todesfällen $H(x)$	Approximations-koeffizient $\alpha(x)$	Nach Formel (5') berechnete Häufigkeit $57\, f(x)$
0	2	1,7	2,7
1	8	1,0	6,3
2	11	0,3	8,8
3	6	1.7	9,4
4	8	0,6	8,4
5	5	3.2	6,8
6	8	— 0,3	5,0
7	2	3,8	3,4
8	3	1,1	2,2
9	2	0,6	1,4
10	1	2,9	0,8
11	1	— 1,9	0,5

In diesem Beispiel erhalten wir für die vier ersten Semiinvarianten, für die Konstanten der Verteilung und für die totalen Kriterien nachstehende Zahlenwerte :

$\mu_1 =$ 4,04
$\mu_2 =$ 6,88
$\mu_3 =$ 11,21
$\mu_4 = -$ 11,02
$\lambda =$ 4,04
$\delta -$ 0,70

$$\frac{\mu_1 \mu_3}{2 \mu_2^2 - \mu_1 \mu_2} = 0{,}68$$

$$\frac{\mu_1^2 \mu_4}{\mu_1^2 \mu_2 - 6 \mu_2^2 \mu_1 + 6 \mu_2^3} = -0{,}20$$

d) Häufigkeit der Todesfälle an Scharlach in der Schweiz in den Monaten des zweiten Halbjahres (Juli bis Dezember) der Jahre 1879 *bis* 1900. — Wie EGGENBERGER (6) gezeigt hat, bilden diese Todesfälle eine Verteilung, die dem Schema der Wahrscheinlichkeitsansteckung unter der Voraussetzung seltener Ereignisse und schwacher Chancenvermehrung gut entspricht. Wir geben die Verteilung der Todesfälle auf die Monate bis zu $x = 30$. Darüber hinaus kamen vor: 36 Fälle in zwei Monaten und in je einem Monat 38, 40, 49 und 57 Fälle.

Die ersten vier Semiinvarianten, die Konstanten der Verteilung und die lokalen Kriterien ergeben folgende Zahlen:

$\mu_1 =$ 11,90
$\mu_2 =$ 103,65
$\mu_3 =$ 1 649,77
$\mu_4 =$ 32 825,27
$\lambda =$ 11,90
$\delta =$ 7,71

$$\frac{\mu_1 \mu_3}{2 \mu_2^2 - \mu_1 \mu_2} = 0{,}97$$

$$\frac{\mu_1^2 \mu_4}{\mu_1^2 \mu_2 - 6 \mu_2^2 \mu_1 + 6 \mu_2^3} = 0{,}78$$

Der Wert für δ weicht von dem bei Eggenberger angegebenen (7,78) ab, ebenso stimmen einige der theoretisch berechneten Häufigkeiten nicht vollständig überein. Die Differenzen sind darin begründet, dass EGGENBERGER bei der Bestimmung von μ_2 durch $n - 1$, wir dagegen durch n dividiert haben.

e) Häufigkeit der Selbstmorde von Knaben im Alter von unter 10 *Jahren in Preussen von* 1869 *bis* 1893. — Dieses Beispiel ist zuerst von V. BORTKIEWICZ (8) behandelt worden, der die Verteilung durch die Poissonsche Verteilung annäherte. Wie GULDBERG ((2), Seite 272) gezeigt hat, eignet sich aber eine Pascalsche Verteilung besser. Da bei der Pascalschen Verteilung $\mu_2 > \mu_1$, kann man ebensogut vom Schema der Wahrscheinlichkeitsansteckung Gebrauch machen, wobei die Wahrscheinlichkeitsansteckung positiv ist.

Zahl der Todesfälle in 1 Monat x	Anzahl der Monate mit x Todesfällen $H(x)$	Approximations-koeffizient $\alpha(x)$	Nach Formel (5′) berechnete Häufigkeit $132\, f(x)$
0	4	0,7	4,8
1	4	2,6	6,4
2	9	1,1	7,2
3	10	0,4	7,5
4	8	2,0	7,6
5	10	— 0,6	7,4
6	6	0,4	7,2
7	5	2,5	6,8
8	6	0,3	6,4
9	5	— 2,9	6,0
10	2	13,6	5,6
11	5	6,9	5,2
12	8	— 1,8	4,8
13	5	1,8	4,5
14	5	— 0,3	4,1
15	4	2,0	3,8
16	4	2,1	3,4
17	4	— 7,7	3,1
18	1	30,0	2,9
19	3	— 12,3	2,6
20	—	∞	2,4
21	1	50,7	2,1
22	4	— 10,0	1,9
23	1	55,2	1,8
24	4	— 6,4	1,6
25	2	2,8	1,4
26	2	— 6,9	1,1
27	1	3,0	1,0
28	1	3,1	0,9
29	1	3,2	0,9
30	1	— 19,4	0,8

GULDBERG erhält folgende Werte:

Zahl der Todesfälle in 1 Jahr x	Anzahl der Jahre mit x Todesfällen $H(x)$	Approximations-koeffizient $\alpha(x)$	Nach Formel (5') berechnete Häufigkeit $25\,f(x)$
0	4	1,3	4,
1	8	0,7	6,2
2	5	0,9	5,5
3	3	0,3	3,6
4	4	0,1	2,
5	—	∞	0,9
6	1	0,8	0,4

Er erhält weiter für μ_1 und μ_2 die Werte

$$\mu_1 = 1{,}96 \qquad \mu_2 = 2{,}46$$

Daraus ist

$$\lambda = 1{,}96 \quad \text{und} \qquad \delta = 0{,}25$$

* * *

Ohne der Wirklichkeit Zwang anzutun, kann man in den fünf behandelten Beispielen annehmen, dass die einzelnen Todesfälle voneinander nicht ganz unabhängig sind, dass also das Schema der Wahrscheinlichkeitsansteckung zugrundegelegt werden darf. Bei den Scharlachtodesfällen liegt diese Tatsache offen zutage. Für die Statistik wäre es eine interessante Aufgabe, bei den unter *a*), *b*) und *c*) genannten Todesursachen die Verhältnisse nach dieser Richtung hin eingehend zu untersuchen.

LITERATURVERZEICHNIS

(1) GULDBERG, ALF: *On discontinuous frequency-functions and statistical series.* « Skandinavisk Aktuarietidskrift », 1931, p. 167-187.

(2) *derselbe*: *Les fonctions de fréquence discontinues et les séries statistiques.* « Annales de l'Institut H. Poincaré », Vol. III, 1933, p. 229-278.

(3) *derselbe*: *On discontinuous frequency-functions of two variables.* « Skandinavisk Aktuarietidskrift », 1934, p. 89-117.

(4) Frisch, Ragnar : *On the use of difference equations in the study of frequency distributions.* « Metron », Vol. X, 1932, p. 35-59.
(5) Pólya, G. : *Sur quelques points de la théorie des probabilités.* « Annales de l'Institut H. Poincaré », Vol. I, 1931, p. 117-161.
(6) Eggenberger, F. : *Die Wahrscheinlichkeitsansteckung. Ein Beitrag zur theoretischen Statistik.* Diss. Zürich, 1924. Erschienen in « Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker », 19. Heft, 1924, S. 31-144.
(7) Eidgenoessisches Statistisches Amt : *Ehe, Geburt und Tod in der schweizerischen Bevölkerung während der Jahre* 1901-1920. « *Schweizerische Statistische Mitteilungen* », X. Jhrg., 1928, 4. Heft, S. 378 und 379.
(8) v. Bortkiewicz, L. : *Das Gesetz der kleinen Zahlen.* Leipzig, 1898. S. 17 und 18.

E. ZWINGGI

Zur Frage des Beharrungszustandes

I.

Sterblichkeit und Fruchtbarkeit bestimmen, abgesehen von den relativ nicht sehr bedeutenden Wanderungen, die Grösse einer Bevölkerung vollständig. Jede zutreffende Vorausberechnung des künftigen Bevölkerungsstandes fusst daher auf einer plausiblen Annahme über den mutmasslichen Verlauf der Sterblichkeit einerseits und der Fruchtbarkeit anderseits.

Die Fortschreibung der Bevölkerungsgrösse auf Grund der Fruchtbarkeits- und Sterblichkeitsverhältnisse ist schon verschiedentlich untersucht worden. In erster Linie sind die Arbeiten von A. J. Lotka (1) zu nennen, der die ganze Materie umfassend behandelt hat. Auch Wicksell (2), Stránsky und Bulina (3) haben zur mathematischen Theorie und praktischen Anwendung dieser Vorausberechnungsart wesentlich beigetragen. Wir möchten in dieser kurzen Arbeit ebenfalls einen Beitrag zu diesem Problem liefern und speziell auf die Frage des *Beharrungszustandes* eintreten.

Je nach der getroffenen Annahme über den künftigen Verlauf der Sterblichkeit und der Fruchtbarkeit wird auch die Wachstumskurve der untersuchten Bevölkerung mehr oder weniger stark ansteigen, u. U. wird sie einem maximalen Wert zustreben und in dieser Lage dann verharren oder aber auch wieder absinken. Die Frage nach den Bedingungen, unter denen eine Bevölkerung gegen eine endliche Grenzlage hinstrebt, stellt sich also ganz von selber. Die gesuchten Bedingungen sind einzig und allein Funktion der Sterblichkeit und der Fruchtbarkeit, die absolute Grösse der Bevölkerung spielt keine Rolle mehr. Variiert die Sterblichkeit, so muss auch die Fruchtbarkeit ändern, damit der Ruhezustand gewahrt bleibt. Die Bestimmung der Variation der

einen Masszahl als Funktion der andern soll hier ausführlich dargelegt werden. Die gefundene exakte mathematische Lösung kann ohne weiteres auch zu praktischen Berechnungen verwendet werden, wie wir am Schlusse unserer allgemeinen Ableitungen zeigen wollen.

II.

Als *Ausgangsgleichung* wählen wir eine Integralgleichung für die Geburtenzahl und spezialisieren sie dann auf die besonderen Verhältnisse im Beharrungszustand. Die Ableitungen werden nur für eine weibliche Bevölkerung durchgeführt, die Uebertragung der Ergebnisse auf eine männliche Bevölkerung bietet keine grossen Schwierigkeiten.

Wir führen die folgenden Bezeichnungen ein:

$B(t)$: Grösse der weiblichen Bevölkerung im Zeitpunkt t.

$n(y)$: Alterszusammensetzungsfunktion der Ausgangsbevölkerung $B(0)$.

$\sigma(\tau, y)$: Intensität der Fruchtbarkeit, gemessen im Zeitpunkt τ, bezogen auf eine y-jährige Frau gleich welchen Zivilstandes. und auf Mädchengeburten.

$p_0^{\tau}(t)$: t-jährige Ueberlebenswahrscheinlichkeit einer o-jährigen Frau, die im Zeitpunkt τ geboren wurde.

$p_y^0(t)$: t-jährige Ueberlebenswahrscheinlichkeit einer y-jährigen Frau, vom Zeitpunkt $\tau = 0$ an gemessen.

μ_y^{τ}: Intensität der Sterblichkeit einer y-jährigen Frau, im Zeitpunkt τ gemessen.

$G(t)\,dt$: Zahl der Mädchengeburten im Zeitintervall t bis $t + dt$.

Ein geschlossener, im Zeitpunkt τ gebildeter Bestand l_y^{τ} von lauter y-jährigen Frauen hat allgemein nach ξ Jahren noch den Umfang $l_{y+\xi}^{\tau}$; im nachfolgenden Zeitintervalle ξ bis $\xi + d\xi$ beträgt die Veränderung

$$d\,l_{y+\xi}^{\tau} = -\,l_{y+\xi}^{\tau} \cdot \mu_{y+\xi}^{\tau+\xi}\,d\xi\,.$$

Die Integration zwischen den Grenzen $\xi = 0$ und $\xi = t$ liefert

$$(1) \qquad {}_t p_y^{\tau} = \frac{l_{y+t}^{\tau}}{l_y^{\tau}} = e^{-\int_0^t \mu_{y+\xi}^{\tau+\xi} \cdot d\xi}$$

Damit sind aber die benötigten Wahrscheinlichkeiten ${}_t p_0^{\tau}$ und ${}_t p_y^0$ durch die Intensität der Sterblichkeit dargestellt; diese Ausdrucksweise wird dann später zur Erfassung der Sterblichkeitsvariation zu verwenden sein.

Die Gesamtzahl der Geburten im Zeitmoment t bis $t + dt$ setzt sich aus zwei Komponenten zusammen. Einmal aus der Zahl der Geburten, die aus dem Eröffnungsbestand $B(0)$ hervorgeht; dann geben aber auch die im Zeitintervall 0 bis t Geborenen selber wieder zu neuen Geburten Anlass. Allgemein t Jahre nach Beginn unserer Betrachtung sind vom Jahrgang $B(0) \cdot n(y) \cdot dy$ noch $B(0) \cdot n(y) \cdot {}_t p_y^0 \cdot dy$ Elemente am Leben. Sie geben aber Veranlassung zu

$$B(0) \cdot n(y) \cdot \sigma(t, y+t) \cdot dy \cdot dt$$

Geburten. Anderseits leben von den $G(\tau) \cdot d\tau$ im Zeitmoment τ bis $\tau + d\tau$ Geborenen im Zeitpunkt t noch $G(\tau) \cdot {}_{t-\tau} p_0^{\tau} \cdot d\tau$; aus diesen Elementen entstehen aber

$$G(\tau) \cdot {}_{t-\tau} p_0^{\tau} \cdot \sigma(t, t-\tau) \cdot d\tau \cdot dt$$

neue Geburten. Summiert man noch über alle Altersstufen, so folgt für die gesamte Zahl der Geburten im Zeitmoment t bis $t + dt$:

$$(2) \qquad G(t) \cdot dt = B(0) \int_0^{\infty} n(y) \cdot {}_t p_y^0 \cdot \sigma(t, y+t) \cdot dy \cdot dt +$$

$$+ \int_0^t G(\tau) \cdot {}_{t-\tau} p_0^{\tau} \cdot \sigma(t, t-\tau) \cdot d\tau \cdot dt .$$

Die Altersverteilungsfunktion $n(y)$ ist durch eine Bestandesaufnahme vollständig zu ermitteln. Die gesuchte Funktion $G(t)$ lässt sich dann aus der abgeleiteten Integralgleichung berechnen, sobald über den Verlauf von ${}_t p_0^{\tau}$ und $\sigma(t, y)$ bestimmte Annahmen getroffen worden sind. Die Beobachtungen weisen darauf hin, dass $\sigma(t, y)$ eine zeitlich für das gleiche Alter y abnehmende

Funktion ist ; dagegen muss $p_0^\tau(t)$ mit zunehmendem τ für gleiches t wachsend vorausgesetzt werden.

Wir haben nun die gewonnene Beziehung (2) auf den Beharrungszustand zu spezialisieren. Einmal ist

$$\left| p_y^0(t) \right|_{t \to \infty} = 0 .$$

Dann setzen wir noch :

$$\sigma(\infty, y) = \sigma(y) ,$$

$$p_0^\infty(t) = p_0(t) ,$$

$$G(\infty) = G = \text{konstant.}$$

Das zweite Integral in Gleichung (2) kann mit diesen Setzungen und der Transformation $t - \tau \to \tau$ auch geschrieben werden als

$$G\,dt = G \int_0^\infty p_0(\tau) \,.\, \sigma(\tau) \,.\, d\tau \,.\, dt \tag{3}$$

oder

$$\text{I} = \int_0^\infty p_0(\tau) \,.\, \sigma(\tau) \,.\, d\tau \,. \tag{4}$$

Diese letzte den stationären Zustand beschreibende Beziehung ist bekannt.

III.

In der eben abgeleiteten Relation (4) wollen wir nun die Fruchtbarkeit $\sigma(\tau)$ variieren lassen und die dazu äquivalente Veränderung von $p_0(\tau)$ zu bestimmen suchen, derart, dass die Gleichung (4) bestehen bleibt. Wir bezeichnen mit τ_0 das tiefste Alter, bei dem $\sigma(\tau)$ gemessen wird. Dann führen wir noch den Begriff der relativen Fruchtbarkeit $\varphi(\tau)$ ein ; $\varphi(\tau)$ ist definiert durch

$$\varphi(\tau) = \frac{\sigma(\tau)}{\sigma(\tau_0)} .$$

Beziehung (4) schreiben wir darauf wie folgt um :

$$\frac{\text{I}}{\sigma(\tau_0)} = k = \int_{\tau_0}^\infty p_0(\tau) \,.\, \varphi(\tau) \,.\, d\tau \,. \tag{5}$$

Nun lassen wir die Fruchtbarkeitsverhältnisse so variieren, dass an der Aenderung alle Altersklassen relativ gleich stark beteiligt sind. In Wirklichkeit trifft diese Voraussetzung nicht genau zu, mangels bestimmter Erkenntnis über das Mass der Veränderung in den einzelnen Altersstufen sind wir aber zu dieser Annahme gezwungen. Die Ausgangsfunktion bezeichnen wir mit $\sigma^0(\tau)$; die variierte sei

$$\sigma(\tau) = \alpha \cdot \sigma^0(\tau) . \tag{6}$$

Dem Werte $\sigma^0(\tau)$ soll immer $\sigma^0(\tau_0)$ koordiniert sein Die. relative Fruchtbarkeit $\varphi(\tau)$ bleibt aber bei allen Variationen der absoluten Fruchtbarkeit $\sigma(\tau)$ unverändert.

Zum Ausgleich der veränderten Fruchtbarkeit lassen wir die Sterblichkeit nun ebenfalls variieren. Auch hier nehmen wir mangels genauer Erfahrung an, dass alle Altersstufen an dieser Aenderung in gleicher Weise beteiligt seien. Die neue Sterblichkeitsintensität $\bar{\mu}_y$ wird aus der alten μ_y gebildet durch

$$\bar{\mu}_y = (1 - \beta) \cdot \mu_y . \tag{7}$$

Analog zu Gleichung (1) können wir für die variierte Ueberlebenswahrscheinlichkeit $\bar{p}_0(\tau)$ schreiben :

$$\bar{p}_0(\tau) = e^{-\int_0^\tau (1-\beta)\mu_\xi \, d_\xi} = e^{-\int_0^\tau \mu_\xi \, d_\xi} \cdot \left\{ e^{-\int_0^\tau \mu_\xi \, d_\xi} \right\}^{-\beta} \tag{8}$$

$$= p_0(\tau) \cdot \{ p_0(\tau) \}^{-\beta} .$$

Indem wir die Ergebnisse (6) und (8) in (5) einsetzen, finden wir :

$$k = \alpha \int_{\tau_0}^{\infty} p_0(\tau) \cdot \varphi(\tau) \cdot \{ p_0(\tau) \}^{-\beta} \cdot d\tau , \tag{9}$$

oder nachdem wir $\{ p_0(\tau) \}^{-\beta}$ in eine Reihe entwickelt haben :

$$k = \alpha \int_{\tau_0}^{\infty} p_0(\tau) \cdot \varphi(\tau) \left\{ 1 + \sum_{i=1}^{\infty} \frac{(-1)^i \cdot \beta^i \cdot \{ l\,n\, p_0(\tau) \}^i}{i!} \right\} d\tau . \tag{10}$$

Wir bezeichnen allgemein mit

$$S_i = \frac{1}{i!} \int_{\tau_0}^{\infty} p_0(\tau) \cdot \varphi(\tau) \cdot \{ l\,n\, p_0(\tau) \}^i \cdot d\tau . \tag{11}$$

Gleichung (10) lautet dann endgültig:

$$k = \alpha \left\{ k + \sum_{i=1}^{\infty} (-1)^i \cdot \beta^i \cdot S_i \right\} \cdot \tag{12}$$

Damit ist zwischen der Variation der Fruchtbarkeit α und der Sterblichkeitsveränderung β ein einfacher Zusammenhang erstellt. Sobald β gegeben ist, kann auch α durch eine einfache Summation bestimmt werden; dabei ist wichtig zu bemerken, dass die Glieder S_i rasch abnehmen, so dass man sich mit einigen wenigen Gliedern begnügen kann. Diese rasche Konvergenz ist natürlich Bedingung, um die Formel auch praktisch anwenden zu können.

Wir gehen nun zu den *praktischen Anwendungen* über. Wenn wir die beiden Grundgrössen $p_0(\tau)$ und $\sigma(\tau)$ in die Beziehung (4) gerade so einführen, wie sie aus den statistischen Erhebungen eines Landes hervorgehen, so wird diese Gleichung nicht erfüllt sein, da diese beiden Werte nicht aus einer stationären Bevölkerung stammen. Man wird daher bloss von der relativen Fruchtbarkeit $\varphi(\tau)$ im heutigen Zeitpunkt ausgehen und dieses $\varphi(\tau)$ sogleich als Ausgangswert festsetzen. Aus der als erfüllt angesehenen Gleichung (5) kann man dann noch $\sigma^0(\tau_0)$ berechnen. Ein Vergleich des so bestimmten $\sigma^0(\tau_0)$ mit dem heutigen $\sigma(\tau_0)$ gibt schon einen ersten Anhaltspunkt dafür, in welcher Richtung sich bei gegebener Sterblichkeit die Fruchtbarkeit entwickeln darf, damit die Bedingung des Beharrungszustandes erhalten bleibt.

Nehmen wir einmal an, dass der aus (5) berechnete Wert $\sigma^0(\tau_0)$ grösser sei als der auf Grund der letzten Erfahrungen bestimmte Wert $\sigma(\tau_0)$ und $\alpha \cdot \sigma^0(\tau_0)$ betrage. Wie stark müsste sich nun die Sterblichkeit verändern, damit mit den heutigen Fruchtbarkeitsverhältnissen der Beharrungszustand gewahrt bleibt? Das Mass der Veränderung β geht sofort aus (12) hervor, wenn wir die aus

$$\tau(\tau_0) = \alpha \cdot \sigma^0(\tau_0)$$

bestimmte Grösse α in diese Gleichung einsetzen und nach β auflösen.

Umgekehrt wollen wir einmal annehmen, dass die noch mögliche Lebensverbesserung β abgeschätzt sei und dass nun bestimmt werden soll, wir gross der entsprechende Rückgang der

Fruchtbarkeit sein darf. Die Lösung dieser Aufgabe ist noch einfacher, man braucht bloss β in (12) einzuführen und nachher nach α aufzulösen.

Die statistischen Berichte geben gewöhnlich die Fruchtbarkeitsverhältnisse nur für Altersgruppen und nicht für einzelne Alter wieder. Diesem Umstande müssen wir Beziehung (12) noch anpassen. Mit n bezeichnen wir die Ordnung der jeweilen 5 Altersjahre umfassenden Gruppe ; die $n+1^{\text{ste}}$ Altersgruppe enthält also die Alter von $5n$ bis $5n+4$. Die relative Furchtbarkeit, die der Gruppe n entspricht, sei $\varphi(n)$. Dann sei noch n_0 die Gruppe die erstmals einen Wert $\varphi(n)$ aufweist. Das allgemeine Glied S_i lässt sich dann aus (11) umformen zu :

$$(13)\quad \overset{*}{S}_i = \frac{1}{i!} \sum_{n=n_0} \sum_{\lambda=0}^{4} p_0(5n+\lambda) \cdot \varphi(n) \cdot \{ l\, n\, p_0(5n+\lambda) \}^i .$$

Mit Hilfe der neuen $\overset{*}{S}_i$ lässt sich α wieder aus

$$(14)\quad k = \alpha \left\{ k + \sum_{i=1}^{\infty} (-1)^i \cdot \beta^i \cdot \overset{*}{S}_i \right\}$$

berechnen. Wir wollen β als gegeben ansehen ; für α finden wir dann

$$(15)\quad \alpha = \frac{k}{k + \sum_{i=1}^{\infty} (-1)^i \cdot \beta^i \cdot \overset{*}{S}_i} .$$

Im allgemeinen wird β nur in engen Grenzen variieren, z. B. von $\beta = 0$ bis $\beta = 0.5$. Um den funktionalen Zusammenhang von α mit β zu erkennen, dürfen wir daher in erster Annäherung alle Glieder in β^2, β^3 usw. weglassen, da zudem auch die Reihe der $\overset{*}{S}_i$ rasch abnimmt. Es bleibt also

$$(16)\quad \overset{*}{\alpha} \sim \frac{k}{k - \beta \overset{*}{S}_1} .$$

Falls man nun diesen neuen Ausdruck in eine Reihe entwickelt und ebenfalls alle Glieder in β^2, β^3 usw. vernachlässigt, so folgt :

$$(17)\quad \overset{*}{\alpha} \sim 1 + \frac{\overset{*}{S}_1}{k} \beta .$$

Damit haben wir das bedeutungsvolle Resultat erhalten, dass die Veränderung der Fruchtbarkeit α eine *lineare Funktion* der Sterblichkeitsvariation β ist.

Wir wollen nun noch ein Beispiel zahlenmässig durchrechnen. Einmal müssen wir zeigen, dass die Reihe der $\tilde{S}_i$ wirklich rasch abnimmt, und dann haben wir auch noch die eben abgeleitete lineare Abhängigkeit nachzuweisen. Als Absterbeordnung im Beharrungszustand nehmen wir die schweizerische Sterbetafel für Frauen nach den Beobachtungen 1920-21. Die relative Furchtbarkeit entnehmen wir den amerikanischen Beobachtungen (4), die Zahlen wurden aber vorher der ungleichen Verteilung der Verheirateten in der Schweiz und in den U.S.A. angepasst. Zuerst bestimmten wir die Reihe der $\tilde{S}_i$:

$$\begin{aligned} k &= 25.0555 \\ \tilde{S}_1 &= -\ 4.9965 \\ \tilde{S}_2 &= 1.0215 \\ \tilde{S}_3 &= -\ 0.2150 \end{aligned}$$

Die Abhängigkeit der Fruchtbarkeitsveränderung α vom Parameter β ist durch die folgende Wertereihe dargestellt :

β	α
0.05	0.9900
0.10	0.9800
0.20	0.9601
0.30	0.9401
0.40	0.9201
0.50	0.9001

Die Werte α wurden aus der Reihe (15) bestimmt, wobei alle Glieder bis β^3 mitgenommen worden sind.

Um die Genauigkeit der Entwicklung (15) zu prüfen, haben wir die Grösse α auch unter Umgehung der Reihenentwicklung berechnet ; wir mussten dazu jeweils die neue Absterbeordnung von Anfang an aufstellen. Zuerst haben wir eine Lebensverschlechterung von 10 % ($\beta = -0.10$) und dann eine Lebensverbesserung von 45 % ($\beta = 0.45$) betrachtet. Wir fanden bei

	$\beta = -0.10$	$\beta = 0.45$
als genauen Wert für α	1.021	0.913
nach Gleichung (15)	1.020	0.910

Auch bei relativ grossen Veränderungen von β genügt die abgeleitete Näherungsformel (15) vollständig.

Die im vorletzten Abschnitte aufgeführte Zahlenreihe lässt die lineare Abhängigkeit klar erkennen ; für unsere Rechnungsvoraussetzungen lautet Beziehung (17) :

(18) $$\overset{\circ}{\alpha} = 1 - 0.1994 \cdot \beta$$

Die nach der Beziehung (18) berechneten Werte für α dürfen im Hinblick auf die für das spezielle Problem notwendige Genauigkeit als vollständig ausreichend bezeichnet werden. Wir fanden

β	α	$\overset{\circ}{\alpha}$	Δ
0.05	0.9900	0.9900	0.0000
0.10	0.9800	0.9801	— 0.0001
0.20	0.9601	0.9601	0.0000
0.30	0.9401	0.9402	— 0.0001
0.40	0.9201	0.9202	— 0.0001
0.50	0.9001	0.9003	— 0.0002

Bemerkenswert ist, dass die Sterblichkeitsveränderung rund 5 mal so gross sein muss als die Variation der Furchtbarkeit. Wenn man bedenkt, dass verschiedene Bevölkerungsvorausberechnungen mit der Annahme eines bis zu 25 %-igen Fruchtbarkeitsrückganges durchgeführt wurden, so kann man leicht ermessen, welche Lebensverbesserung zur Aufrechterhaltung eines stationären Zustandes erforderlich ist ; u. U. erfolgt dann überhaupt keine Stabilisierung mehr auf einen festen Grenzwert.

Obwohl wir das Problem des Zusammenhanges zwischen Sterblichkeit und Fruchtbarkeit unter einschränkenden Voraussetzungen behandeln mussten, scheinen uns die Ergebnisse nicht ohne Bedeutung zu sein. Jede Bevölkerungsvorausberechnung muss schliesslich auch den Endzustand enthalten, und zur Bestimmung dieser Grenzlage können die abgeleiteten Beziehungen gute Dienste leisten.

LITERATUR

(1) A. J. Lotka : Wir nennen hier nur einige neuere Arbeiten, die im Hinblick auf die vorliegende Untersuchung von Interesse sind.

a) *On the true rate of natural increase. As exemplified by the population of the United States.* « Journal of the American Statistical Association », 1925.

b) *The measure of net fertility.* « Journal of the Washington Academy of Sciences », Vol. 15, 1925.

c) *The progressive adjustment of age distribution to fecundity.* « Journal of the Washington Academy of Sciences », Vol. 16, 1926.

d) *The progeny of a population element.* « The American Journal of Hygiene », Vol. 8, 1928.

e) *The spread of generations.* « Human Biology », Vol. 1, 1929.

f) *The true rate of natural increase of the population of the United States.* « Metron », Vol. 8, 1930.

g) *Applications de l'analyse au phénomène démographique.* « Journal de la Société de Statistique de Paris », 1933.

(2) S. D. Wicksell : *Nuptiality, fertility, and reproductivity.* « Skandinavisk Aktuarietidskrift », Jahrgang 14, 1931.

(3) J. Stránsky & J. Bulina : *The development of population in the Czechoslovak Republic.* « Aktuárské Vědy », 1. Band, 1930.

(4) A. J. Lotka : *The size of American families in the eighteenth century.* « Journal of the American Statistical Association », 1927.

V. CASTELLANO

Recente letteratura sugli indici di variabilità

Alcuni recenti lavori (1) hanno richiamato l'attenzione degli statistici sulle questioni della preferenza da accordarsi ai vari indici di variabilità proposti, e contrapposti talvolta l'un l'altro, dai vari Autori, e delle relazioni che intercedono fra essi.

Primo tra questi, in ordine di tempo, un lavoro dello Yntema che, sul «Giornale dell'Associazione Statistica Americana», mostra la confusione che deriva dall'uso di molteplici indici per chi prescinda dalle specifiche ragioni di ciascuno. Secondo, un lavoro del Mortara, che, in una comunicazione alla Sessione di Londra dell'Istituto Internazionale di Statistica, tenta di dimostrare la legittimità, come misura della variabilità, di due soli di tali indici: la differenza media e lo scostamento semplice medio, ma la opportunità di uno solo di questi: lo scostamento semplice medio. Terzo, un altro lavoro presentato dal Pietra alla stessa riunione dell'Istituto Internazionale di Statistica, nel quale questo A. precisa e completa i risultati già da lui raggiunti molti anni addietro sull'argomento della discordanza fra alcuni indici di variabilità. Quarto, un lavoro del D'Addario (che ha provocato, insieme al

(1) D. B. Yntema: *Measures of the inequality in the personal distribution of wealth or income.* « Journal of the American Statistical Association », Vol. XXVIII, IV, 184, December, 1933.

G. Mortara: *Sulle disuguaglianze statistiche.* « XXII Sessione dell'Istituto Internazionale di Statistica », Londra, 1934.

G. Pietra: *Intorno alla discordanza tra gli indici di variabilità e di concentrazione.* « XXII Sessione dell'Istituto Internazionale di Statistica », Londra, 1934.

R. D'Addario: *Sulla misura della concentrazione dei redditi.* « Istituto Poligrafico dello Stato », Roma, 1934.

V. Dominedò: *Sulla differenza media.* « Giornale degli Economisti e Rivista di Statistica », Novembre 1934.

precedente, una replica del Pietra), nel quale si afferma, limitatamente alle leggi analitiche di distribuzione dei redditi sin qui proposte, la perfetta concordanza del rapporto di concentrazione con lo scostamento semplice medio e la identità tra gli indici α del Pareto e δ del Gini. Ultimo, uno studio del Dominedò nel quale è messo in luce ancora un altro significato del rapporto di concentrazione e la impossibilità di sostituire, per tutti gli scopi, a questo indice lo scostamento semplice medio.

Data l'importanza che l'argomento riveste, sia dal punto di vista teorico, sia dal punto di vista delle applicazioni — poichè non è dubbio che il capitolo della variabilità sia un capitolo fondamentale della Metodologia e che la misura della variabilità dei caratteri abbia gran parte nella statistica applicata — e data la divergenza delle opinioni espresse in proposito da alcuni Autori, non mi sembra privo di utilità un esame analitico delle principali questioni in argomento trattate dagli Autori citati.

Per comodità del lettore, distinguo in tre gruppi tali questioni :

1° Relazione tra gli indici α del Pareto e δ del Gini. Natura e limiti di questa relazione.

2° Relazioni tra lo scostamento semplice medio S e il rapporto di concentrazione R, o differenza media relativa (1), e precisamente :

a) se gli indici S e R sono concordi o discordi ;

b) se gli indici S e R dipendono dalla media aritmetica ;

c) se R possa fornire approssimazioni comparabili a quelle fornite da S quando i dati sono noti solo per classi.

3° Se esiste una « misura » della variabilità. È in questa ultima questione che trova il suo posto naturale la domanda postasi da qualche Autore, se uno dei due indici S o R sia preferibile all'altro.

---

(1) Qui e nel seguito, indipendentemente dalle notazioni, sovente discordanti, dei varii Autori, indico con S lo scostamento semplice medio dalla media aritmetica, sia assoluto che relativo, con R il rapporto di concentrazione e con Δ la differenza media, sia assoluta che relativa. Quando faccio particolare riferimento agli indici relativi, adotto le notazioni $\frac{S}{2A}$ e $\frac{\Delta}{2A}$. Quando la semplice notazione S potesse dar luogo ad equivoci e quindi, in particolare, quando considero, insieme allo scostamento semplice medio dalla media aritmetica anche quello dalla mediana — che indico con ${}_1S_M$ in generale e con $\frac{{}_1S_M}{A}$ quando mi riferisco in particolare all'indice relativo — sostituisco alla notazione S l'altra, più completa, di ${}_1S_A$.

Relazione tra α e δ.

Prima questione è quella della natura della nota relazione tra α e δ cui il Mortara (1), e dopo di lui il Bresciani (2) e quindi anche il Ricci (3), pare che attribuissero un carattere rigoroso e necessario, sì che sembra quasi autorizzato da questi il D'Addario a ripresentarcela oggi come una perfetta identità. Con quanta ragione ciò possa farsi, vedremo.

Quali sono, anzitutto, i significati di δ e di α?

δ ha due significati distinti. Se diciamo col Gini (4) che un carattere è tanto più concentrato quanto maggiore è la disuguaglianza tra la frazione $1 - q_i$ dell'ammontare complessivo posseduto da un certo numero degli elementi maggiormente dotati della distribuzione, e la frazione $1 - p_i$ da questi rappresentata sul totale degli elementi, ha significato di « indice di concentrazione » l'esponente δ_i al quale bisogna elevare ciascuna di tali frazioni dell'ammontare del carattere per ottenere la corrispondente frazione degli elementi più dotati, tra i quali detta quantità di carattere è ripartita: $(1 - q_i)^{\delta_i} = (1 - p_i)$. Come media di tali esponenti, δ è un « indice di concentrazione medio » di preciso significato, indipendente da ogni legge analitica cui si possa uniformare la distribuzione, e non lo tocca particolarmente l'osservazione del Mortara (5) di essere maggiore di alcuni, minore di altri δ_i, osservazione che rileva non un difetto, ma una proprietà di tutte le medie (6). Quando i singoli δ_i sono press'a poco uguali, sì che uno

---

(1) Mortara: *Note di economia induttiva: Sulla distribuzione dei redditi.* « Giornale degli Economisti e Rivista di Statistica », maggio 1911.

(2) C. Bresciani-Turroni: *La ricchezza delle città.* « Annali del Seminario Giuridico della R. Università di Palermo », 1913.

(3) U. Ricci: *L'indice di variabilità e la curva dei redditi.* Athenaeum, Roma, 1916.

(4) Cfr. C. Gini: *Indici di concentrazione e di dipendenza.* « Biblioteca dell'Economista ». Serie V, Vol. XX, 1910 e cfr. anche il lavoro dello stesso A.: *Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza,* in « Giornale degli Economisti », gennaio 1909, nel quale è esposta l'idea fondamentale sviluppata nel precedente.

(5) G. Mortara: *Note di economia induttiva: Sulla distribuzione dei redditi,* già cit.

(6) Nel fervore della polemica accesa intorno alla memoria « *Die Disparitätsmasse in der Einkommensstatistik* », presentata dal Bortkiewicz alla XIX Sessione dell'Istituto Internazionale di Statistica, nella quale apparivano come nuovi molti risultati ottenuti dal Gini in *Variabilità e Mutabilità* (Istituto Economico-Giuridico della R. Università di

stesso valore medio δ possa sostituirsi ad essi senza troppo grande errore, allora δ ha anche valore di « indice descrittivo » della distribuzione. Supposta questa continua, δ è il parametro della curva di concentrazione che interpola i dati stessi.

α ha il solo significato di parametro della curva di Pareto e si presenta, quale indice della disuguaglianza della distribuzione, di interpretazione almeno non immediata, se ha potuto errare al riguardo lo stesso Pareto (affermando che la disuguaglianza cresce o diminuisce al crescere o al diminuire di α) e se solo la relazione inversa con δ ha rappresentato la prova inoppugnabile della esattezza della contraria opinione del Benini (1). Il Gini, cui si deve tale relazione, ridimostrata poi, con più o meno vari procedimenti, da più di un Autore (2), avvertiva subito (3) che essa ha un carat-

---

Cagliari, Bologna, Cuppini, 1912), accadde anche al Bortkiewicz di esprimere, a proposito della differenza media, una opinione analoga a quella del Mortara, subito rilevata dal Gini, (Cfr. la risposta del Bortkiewicz alle osservazioni fatte dal Gini alla sua comunicazione e la replica del Gini).

In tale polemica ritiene di dover prendere posizione oggi l'Anderson (*Einführung in die mathematische Statistik*, Wien, 1935, pag. 160) dato che « bisher kein einziger Deutscher Wissenschaftler seine Stimme zum Schutze Bortkiewicz' erhoben hat ». Conveniamo con l'Anderson che è « nicht verwunderlich dass er (Bortkiewicz) ein Thema behandelnd, welches Gini bereits bearbeitet hat, auch zu ähnlichen Resultaten und zu einigen identischen Formeln gelangt ist », ma ci pare di dovere aggiungere anche che ciò è tanto meno sorprendente in quanto è incontroverbile che tali risultati e tali formule già conosceva il Bortkiewicz attraverso Czuber, Weinberg, Dalton e Gumbel, se non anche attraverso Lexis, Bowley, Julin, e March, i quali tutti li attribuiscono debitamente al Gini.

(1) È noto che il Pareto riteneva che la disuguaglianza della distribuzione aumentasse con α, mentre il Benini riteneva il contrario. Indipendentemente da δ si sarebbe potuto risolvere brillantemente la questione, in favore del Benini, se si fosse ricorso al valore di α che corrisponde alla distribuzione uniforme. Questo è evidentemente $\alpha = \infty$ poichè solo per tale valore di α si ricava, con considerazioni di limite, dalla legge di Pareto $N_x = \frac{K}{x^\alpha}$, posta sotto la forma $N_x = N_h \left(\frac{h}{x}\right)^\alpha$: $N_x = N_h$ per $x = h$, $N_x = O$ per $x > h$.

(2) Cfr. C. Gini: *Indici di concentrazione e di dipendenza*, già cit.; V. Furlan: *Neue Literatur zur Einkommens Verteilung in Italien.* « Jarbücher für Nationalökonomie und Statistik », August 1911; G. Mortara: *Note di economia induttiva*, già cit. ; F. Savorgnan: *Intorno all'approssimazione di alcuni indici della distribuzione dei redditi.* « Atti del R. Istituto Veneto di Scienze, Lettere ed Arti », Tomo LXXIV, parte 2ª, 1915.

(3) Cfr. C. Gini: *Indici di concentrazione e di dipendenza*, già cit. ; e *Variabilità e Mutabilità*, già cit.

tere solo teorico e un valore solo approssimativo. Essa significa che, se una distribuzione segue *esattamente* la curva di Pareto e se, essendo $\alpha > 1$, è il reddito massimo così grande rispetto al reddito minimo da potersi considerare praticamente infinito, allora la curva di concentrazione della distribuzione è una curva « δ » del Gini di parametro $\delta = \frac{\alpha}{\alpha - 1}$; e viceversa: se una distribuzione segue *esattamente* la curva di concentrazione δ del Gini, allora essa segue anche la corrispondente curva α del Pareto, che è definita dalla relazione $\alpha = \frac{\delta}{\delta - 1}$. Ma quali distribuzioni reali seguono esattamente le leggi analitiche espresse da queste due curve? Occorre scambiare la distribuzione fittizia rappresentata da due corrispondenti di queste curve con quella effettiva, che non si adatta perfettamente bene nè all'una nè all'altra, per affermare, come fa il D'Addario, che tra α e δ sussista « proprio una identità ». Con ciò egli scambia addirittura la statistica per matematica, e non pensa che appunto perchè le distribuzioni reali non si adattano mai perfettamente bene alle curve teoriche accade che, tra le curve di un dato tipo che rappresentano un certo aspetto delle distribuzioni statistiche e le loro trasformate funzionali che ne rappresentano un altro aspetto, quelle che rappresentano meglio una certa distribuzione non sono in generale corrispondenti (1). Precisamente come non si corrispondono il valore più probabile di una grandezza e il valore più probabile del suo quadrato. Di qui un problema di scelta tra α e δ, e quindi tra le corrispondenti curve di Pareto e di Gini.

In linea di principio che una questione di scelta tra α e δ esista non si può assolutamente negare, come ha fatto invece sempre

---

(1) Pare qui naturale chiedersi perchè pure essendo convinto, come pare, che una curva e una sua trasformata funzionale siano in statistica la stessa cosa (pag. 15) ed essendo convinto che la curva proposta da Amoroso per la rappresentazione della distribuzione dei redditi fornisca agli statistici un potente strumento di analisi (pag. 18), l'A. proponga una nuova curva di concentrazione dei redditi e di questa faccia numerose applicazioni, mentre dovrebbe ritenere più proficuo applicare la curva di concentrazione corrispondente alla curva di Amoroso, la quale curva di concentrazione, invero, almeno nell'esempio, considerato dall'Amoroso nella sua memoria « *Ricerche intorno alla curva dei redditi* » (in « Annali di Matematica pura ed applicata », serie IV, Tomo II, febbraio 1925) per il quale ho eseguito il calcolo, fornisce risultati molto soddisfacenti sotto tutti i punti di vista. Non

il D'Addario, anche quando era di una opinione opposta alla sua attuale, e cioè che α e δ fossero due cose diverse (1).

In pratica le ricerche, eseguite su vasto materiale statistico, da molti autori (2), hanno dimostrato che tale questione esiste, e che va risolta in favore di δ, in quanto tale curva di concentrazione descrive meglio, in generale, le distribuzioni empiriche che non la curva di Pareto. È solo su questo terreno, strettamente statistico, che la questione potrebbe essere riaperta. Per affermare che α e δ sono la stessa cosa, si dovrebbe riprendere in esame il materiale elaborato da tali Autori, ed altro ancora, e mostrare che interpolazioni più soddisfacenti di quelle eseguite da questi conducono a valori dei parametri α e δ per i quali valga precisamente la nota relazione teorica. Ciò avrebbe il vantaggio di assicurare gli statistici della incondizionata validità in statistica di al-

---

altrettanto soddisfacenti mi sembrano le applicazioni della propria equazione fornite dal D'Addario, in quanto rilevo, negli esempi da lui riportati per i quali è indicato l'ammontare del reddito minimo, i seguenti valori, empirici e teorici, del rapporto A/h :

Tav.	I	II	III	IV	V	VI	VII	VIII
$A:h$ valore teorico	2.303	2.225	2.355	3.502	3.962	2.289	3.198	2.775
» » empirico	3.362	3.332	3.392	9.470	11.710	5.045	4.864	3.959

Differenze tutte notevoli, alcune enormi, le quali significano che le corrispondenti curve di frequenza teoriche a parità di reddito medio presenterebbero redditi minimi diversissimi da quelli delle distribuzioni empiriche, o, a parità di redditi minimi, redditi medî diversissimi. Dal che non è difficile arguire che tali curve di frequenza non dovrebbero adattarsi molto bene ai dati empirici. D'accordo con l'A. nell'opinione che spesso le cattive interpolazioni non dipendono dai metodi, ma dall'uso che dei metodi si fa, credo gli si possa concedere che, in parte, le divergenze notate sieno ineliminabili, il che potrebbè ascriversi o a deficienza della curva interpolata o a deficienza della sua trasformata o a deficienza di entrambe, come dovrebbe, per coerenza, ammettere l'A.

(1) Scriveva allora il D'Addario: « Si è andata creando (infatti) una tale confusione di idee e si continua ad assistere ad una monotona lotta tra α e δ, che pensiamo sia il caso di intenderci una volta per sempre.

Quella del Pareto non è l'equazione di una curva di distribuzione ?

E quella del Gini non è l'equazione di una curva di concentrazione ?

Perchè confondere allora l'una con l'altra ? »

(Cfr. R. D'Addario: *Un'indagine sulla dinamica distributiva.* « Rivista Italiana di Statistica », Anno II, n. 3, 1930, pagg. 280-282).

(2) Cfr. C. Gini: *Indici di concentrazione e di dipendenza,* già cit. ; *Variabilità e mutabilità,* già cit. ; F. Corridore: *L'evasione dei redditi,* Roma, Leoscher, 1910; *Relazioni tra affitto reale e valore locativo fiscale nel Belgio,* Roma, Loescher & C., 1911 ; E. Porru: *La concentrazione della ric-*

meno tale relazione teorica. Non credo che si possa approvare l'avventata fiducia che il D'Addario dimostra in argomento. Tanto più che appunto su tale fiducia è fondata l'accusa di essere caduti in « truismi », che il D'Addario aveva lanciato a vari Autori in un suo precedente lavoro (1), ed ora ha ripetuto. L'A. giudica « truismo » la ricerca, eseguita da quegli autori, della concordanza o meno della graduatoria dei valori degli α interpolati con quella dei corrispondenti valori rilevati del reddito medio. Con una tale accusa egli mostra solo di non approvare la diligente prudenza dello statistico avveduto, che, ammesse delle ipotesi, per vie diverse e concordanti le verifica.

La idea della identità di α e δ ha anche indotto l'A. a fare all'Yntema una osservazione sul campo di variabilità di α, che è errata quanto mai. Mentre Yntema, nel lavoro innanzi citato, assegna ad α il campo di variabilità $(0,\infty)$, il D'Addario crede che α debba essere > 1, e si meraviglia che alcuni autori abbiano « persino discusso e costruito su valori minori dell'unità ». Tale meraviglia io credo provenga dal fatto che l'A. non ha riflettuto sulla circostanza che la condizione $\alpha > 1$ è imposta unicamente dall'uso di considerare, nelle distribuzioni dei redditi, in generale infinito il reddito massimo. Tale ipotesi è comoda, generalmente accettabile e pratica, ma certo irreale e non necessaria sicchè quando s'incontrino valori di $\alpha < 1$ non bisognerà rigettare a priori tali valori, ma lasciare anzitutto cadere quella ipotesi (2).

---

chezza delle diverse regioni d'Italia. « Studi economico-giuridici pubblicati per cura della Facoltà di Giurisprudenza della R. Università di Cagliari », Vol. IV, parte I, 1911-12 ; E. Czuber : *Beitrag zur Theorie der statistischer Reihen,* « Versicherungwissenschaftlichen Mitteilungen » Neue Folge, 9 Band, 2 Heft, Wien, 1914; F. Savorgnan: *La distribuzione dei redditi nelle provincie e nelle grandi città dell'Austria.* « Pubblicazioni del Museo Commerciale di Trieste », Trieste, 1912 ; *Intorno all'approssimazione di alcuni indici della distribuzione dei redditi.* « Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti », Tomo LXXIV, parte 2ª, 1915 ; M. Saibante : *La concentrazione della popolazione.* « Metron », Vol. VII, n. 2, 1928 ; S. Orlandi : *Su la distribuzione dei redditi mobiliari in Italia.* « Pubblicazioni dell'Istituto di Statistica della R. Università di Roma », Serie 1ª, Vol. I, Roma, 1933.

(1) Cfr. R. D'Addario: *Intorno alla validità di due teoremi paretiani sulla dinamica distributiva.* « Atti dell'Istituto Nazionale delle Assicurazioni », Vol. VI, Roma, 1934.

(2) All'A. è evidentemente sfuggito che lo stesso Pareto prima di considerare infinito il reddito massimo, pone esplicitamente la condizione

Se ho da esprimere il mio parere la questione tra α e δ è duplice.

Esiste una questione tra α e δ quali indici descrittivi di una distribuzione, inoppugnabilmente risolta oggi in favore di δ, per le migliori approssimazioni che la curva « δ » consente rispetto alla « α ». Tale questione rientra nel campo delle complesse relazioni intercedenti tra le trasformate funzionali delle varie curve rappresentative di una distribuzione statistica, relazioni che potrebbero essere l'oggetto di un capitolo affatto moderno, e credo interessante, della teoria delle interpolazioni statistiche e che esorbita pertanto dai limiti di questa rassegna.

Esiste, d'altra parte, una questione tra α e δ quali indici di variabilità che dovrebbe risolversi avendo riguardo al significato di ciascuno: α non ha, che io sappia, una interpretazione diversa da quella che gli attribuisce la legge di Pareto; il suo significato non è d'immediata comprensione, sì che esso non autorizza conclusioni sicure che quando la legge dalla quale discende valga senza riserve; ma in questo caso, essenzialmente teorico, tutti gli indici sarebbero equivalenti, nel senso che tra tutti gli indici passerebbero relazioni costanti, sicchè sarebbe assolutamente indifferente in pratica l'adozione di un indice piuttosto di un altro. Invece δ, come indice medio di concentrazione, ha il vantaggio di un chiaro significato, che ne giustifica l'adozione laddove questo significato interessa.

Tuttavia come indice medio di concentrazione a δ sembra oggi spesso preferibile (1) un altro indice, il rapporto di concentrazione, nel quale convergono moltissimi significati.

Relazione tra S e R.

Quali sono, anzitutto, i significati di S e di R?

Discende immediatamente dalla definizione di S — media aritmetica dei valori assoluti degli scostamenti dalla media arit-

---

$\alpha > 1$ (cfr. Pareto: *Cours d'économie politique*. Tome II, Lausanne, 1897, nota (961), pagg. 312-13).

(1) Tale fu anche l'opinione del suo Autore quando ebbe concepito il rapporto di concentrazione e le sue relazioni con la curva di concentrazione e con la differenza media (cfr. al riguardo: C. Gini: *Sulla misura della concentrazione e della variabilità dei caratteri*. « Atti del R. Istituto Veneto di Scienze, Lettere ed Arti », Tomo LXXIII, parte 2°, 1913-14, paragrafo 1°).

metica — che S può intendersi anche come il doppio della quantità di carattere posseduta in eccesso, rispetto ad una distribuzione uniforme del carattere, dalle unità nelle quali il carattere ha intensità, superiore alla media, ragguagliato al numero dei casi.

I significati di R sono : indice di concentrazione medio ; doppio dell'area di concentrazione ; differenza media relativa ; scostamento semplice medio *totale*.

Come indice di concentrazione medio, R è il coefficiente per il quale bisogna moltiplicare una certa frazione dei casi, più o meno dotati del carattere, per ottenere il valore probabile della differenza tra tale frazione e quella corrispondente dell'ammontare del carattere, e si ottiene come media ponderata degli indici di disuguaglianza $R_i = \frac{p_i - q_i}{p_i}$, presi con pesi p_i o $R'_i = \frac{(1 - q_i) - (1 - p_i)}{1 - p_i}$ con pesi $1 - p_i$, i quali discendono dallo stesso concetto di concentrazione che è alla base della definizione dell'indice δ.

Come doppio dell'area di concentrazione, R esprime la quantità di carattere che deve essere complessivamente trasferita da ciascun elemento all'elemento che lo precede nella graduatoria crescente delle intensità del carattere, per ottenere la distribuzione uniforme (1).

Come differenza media relativa e come scostamento semplice medio totale, R non può farsi più discendere dal concetto di concentrazione — l'accumularsi del carattere nei casi nei quali il carattere stesso supera un certo limite — ma bensì sempre dal concetto di variabilità — attitudine ad assumere varie modalità quantitative — : nella prima accezione esso è la metà del valore probabile della differenza tra due elementi della distribuzione ragguagliata alla media aritmetica di essa ; nella seconda esso è : *a*) la media aritmetica semplice degli scostamenti semplici medi della serie da ciascun termine (2) ; *b*) una media ponderata degli scostamenti semplici medi ${}^1S_A$ dalla media aritmetica : della serie

---

(1) Cfr. V. Castellano : *Sulla interpretazione dinamica del rapporto di concentrazione*. « Giornale dell'Istituto Italiano degli Attuari », Anno IV, n. 2, aprile 1933.

(2) Cfr. C. Gini : *Variabilità e Mutabilità*, già cit., *Di una estensione del concetto di scostamento medio e di alcune applicazioni alla misura dei caratteri qualitativi*. « Atti del R. Istituto Veneto di Scienze Lettere ed Arti », 1917-18, Tomo LXXVII, Parte II, pagg. 397-461.

data, di ciascuna delle due serie nelle quali questa è divisa dalla media aritmetica, di ciascuna delle due serie nelle quali queste sono a loro volta divise dalla propria media aritmetica, e così via (1) ; *c*) una media ponderata degli scostamenti semplici medi ${}^1S_m$ dalla quantità minima : della serie data e di ciascuna delle serie di $n-1$, $n-2$, . . . 1 solo termine che si ottengono da quella togliendo via via la 1^a, la 2^a e la $n-1$ quantità (2).

Naturalmente tali significati sono tutt'altro che indipendenti, ma tanto invece strettamente collegati l'un l'altro che molti Autori non hanno fatto tra essi alcuna distinzione, e hanno lasciato la loro attenzione polarizzarsi intorno alle questioni sollevate, come nuovo indice di variabilità, dalla differenza media. Prima di seguirli su questa strada esamineremo brevemente le tre questioni sopra richiamate, che rivestono un interesse metodologico a sè, indipendentemente dagli scopi per i quali sono considerate da questi A., escluso solo il Pietra.

a) *Discordanza tra R e S*

È questo della discordanza fra gli indici di variabilità oramai un vecchio argomento che, segnalato e discusso a fondo per primo dal Gini in *Variabilità e Mutabilità*, ha richiamato l'attenzione di quanti Autori hanno applicato insieme allo scostamento semplice medio la differenza media, per lo studio della variabilità dei caratteri.

Studiando il problema dal punto di vista teorico, il Pietra (3) ebbe a determinare già nel 1916 i limiti entro i quali possono variare i valori, ragguagliati ai rispettivi massimi, della differenza media, e degli scostamenti semplici medi dalla media aritmetica e dalla mediana, corrispondenti ad un valore determinato, fissato ad arbitrio nel comune campo di variabilità (0,1), di ciascuno di essi. Di tali determinazioni ora si è valso, per calcolare una misura della

---

(1) Cfr. V. Dominedò : *Sulla differenza media*, già cit.

(2) Cfr. C. Gini : *Variabilità e Mutabilità*, già cit., pag. 36, formu.e (26) e (27) : la interpretazione data non è che la « lettura » di queste formule.

(3) Cfr. G. Pietra : *Delle relazioni tra gli indici di variabilità*. « Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti », Anno 1914-15, Tomo LXXIV, Parte II, Note I e II.

probabilità che due coppie di valori di tali indici sieno discordi, cioè una misura della probabilità che tra due valori di uno stesso indice passi una relazione di grandezza opposta a quella che passa tra i valori corrispondenti di un altro indice (1).

Fra gli indici S e R egli determina una probabilità media di

---

(1) Il calcolo è fondato su due ipotesi: 1ª, che siano ugualmente probabili tutti i valori (compresi tra 0 e 1) che ciascun indice può presentare; 2ª, che presentando un indice un certo valore siano ugualmente probabili tutti i valori possibili per un altro indice. Ammesse tali ipotesi è evidente che la probabilità che si abbia discordanza tra due indici, ad esempio, tra lo scostamento semplice medio della media aritmetica S ed il rapporto di concentrazione R, può calcolarsi in due maniere diverse : o partendo da due valori di S per determinare quali tra i valori corrispondenti di R danno luogo a discordanze ; o partendo da due valori di R per determinare quali tra i valori corrispondenti di S danno luogo a discordanze. Nè i due metodi possono dirsi equivalenti in quanto la legge di probabilità costante che si suppone per l'indice dal quale si parte, e quella ugualmente costante, che si suppone per i valori dell'altro corrispondenti ad un valore determinato del primo, implicano, per il secondo, una legge di probabilità niente affatto costante e quindi diversa da quella che si assumerebbe per questo qualora si partisse, per il calcolo, da esso.

Infatti, se $y = x\,(1 - x)$ è l'equazione della parabola le ordinate dei cui punti rappresentano l'ampiezza del campo di variazione della differenza $R - S$ per ogni valore di S, i valori di R compresi tra x e $x + dx$ sono tutti e soli quelli le cui differenze dagli S cui corrispondono sono comprese tra le parallele alla bisettrice del 2º quadrante condotte per gli estremi dell'intervallino $d\,x$. Di questi, a quelli che appartengono anche alla parallela all'asse y condotta per il punto $x - k\,[o < k < \sqrt{1 - x}\,(1 - \sqrt{1 - x})]$, spetta la probabilità $\dfrac{d\,x}{(x - k)\,(1 - x + k)}$, e quindi se $d\,k$ è la probabilità che il valore di S cada dentro un intervallo qualunque di ampiezza $d\,k$ entro il campo (0 , 1), a tutti spetta la probabilità :

$$dx \int_0^{\sqrt{1-x}\,(1-\sqrt{1-x})} \frac{d\,k}{(x - k)\,(1 - x + k)} = dx \text{ lg. } \frac{x}{(1 - \sqrt{1 - x})\,\sqrt{1 - x}}.$$

Similmente, partendo da R, le ipotesi del testo assegnano ai valori di S compresi tra x e $x + dx$ una probabilità di $dx\ 2$ lg. $\dfrac{x}{1 - \sqrt{1 - x}}$. Ognuno può vedere come varino con x queste due leggi di probabilità. Noi osserviamo solo che la prima assegna ai valori di $R < \dfrac{1}{2}$ una probabilità = 0,3870, mentre la seconda assegna agli stessi valori di S una probabilità = 0,6216.

discordanza di $\frac{3}{20}$, e fra S_m e R o S, una probabilità media di $\frac{1}{5}$; ma, per intendere il valore indicativo di tali cifre, occorre considerare che egli esclude dalla massa dei casi possibili tutti quei casi nei quali la differenza tra i due valori di uno stesso indice è così considerevole che la coppia dei valori corrispondenti dell'altro è certamente concordante con la prima. Se si considerano anche tali casi di concordanza, esclusi dal Pietra, i valori di tali probabilità

Partendo da S e da R si ottiene per S_m una stessa legge che attribuisce una probabilità costante, $\lg 2\, dx$ ai valori compresi tra x e $x + dx$ quando $x < \frac{1}{2}$, e variabile, $\lg \frac{1}{2x(1-x)}\, dx$, per $x > \frac{1}{2}$. Questa legge assegna ai valori di $S_m > \frac{1}{2}$ una probabilità complessiva = 0,3456. Viceversa, partendo da S_m si ottiene per S e per R una stessa legge che assegna ai valori x di questi $< \frac{1}{2}$ una densità di probabilità costante $= 2 \lg 2$ ed ai valori $x > \frac{1}{2}$ una densità variabile uguale a $\lg \frac{1}{x}$. Questa legge assegna complessivamente ai valori dei due indici $< \frac{1}{2}$ la probabilità 0,6932. La uguaglianza dei risultati cui conducono i due metodi nel caso di S e R costituisce una prova a posteriori della opportunità delle ipotesi. Non è però certo che i due metodi non conducano agli stessi risultati anche nel caso di S_m e R o S, sebbene l'A. giunga per S_m e S a risultati diversi. Ciò potrebbe dipendere solo dal fatto che il Pietra non ha tenuto conto della circostanza che, data una coppia di valori: $S = x$ e $S' = x + k$, dei quali, posto $k > 0$, si suppone il 1° minore del secondo, quando $x < \frac{1}{2}$ la probabilità che la coppia corrispondente dei valori S_m e S'_m (e analogamente si direbbe per R e R'), presenti una discordanza, non ha la stessa espressione analitica se $x + k \lesseqgtr \frac{1}{2}$. Infatti dato $S = x\left(x < \frac{1}{2}\right)$ e $S' = x + k$, i casi possibili per S_m sono sempre $x\ (x \leq S_m \leq 2x)$ e i casi possibili per S'_m sono $x + k$ se $x + k < \frac{1}{2}$ $(x + k \leq S'_m \leq 2(x + k)$ e $1 - x - k$ se $x + k > \frac{1}{2}$ $(x + k \leq S_m \leq 1)$. (Cfr. G. Pietra, nota citata).

Tenendo conto di questa circostanza si vede facilmente che il 1° integrale che sta a secondo membro della sua formula (25) deve essere spezzato in tre i cui limiti di integrazione sono, per x e k rispettivamente, 0, $\frac{1}{4}$ e 0, x nel 1°, $\frac{1}{4}$, $\frac{1}{2}$ e 0, $\frac{1}{2} - x$ nel 2°, $\frac{1}{4}$, $\frac{1}{2}$ e $\frac{1}{2} - x$, x nel 3°,

diventano $\frac{1}{20}$ la prima e $\frac{1}{10}$ la seconda (1), valori, questi, pur sempre notevoli, sebbene inferiori a quelli dati dal Pietra.

Per contro il D'Addario afferma che, almeno relativamente alle distribuzioni che seguono esattamente le leggi espresse dalle equazioni paretiane di prima e di seconda approssimazione, dall'equazione di Amoroso, da quella riproposta dal Gibrat e da quella

---

essendo anche sostituito in quest'ultimo caso, al fattore $x + k$, che compare al denominatore della funzione integranda, il fattore $1 - x - k$.

La formula (25) del Pietra diventa quindi:

$$P_4 = \frac{1}{2}\int_0^{\frac{1}{4}} dx \int_0^{x} \frac{(x-k)^2}{x\,(x+k)}\,dx + \int_{\frac{1}{4}}^{\frac{1}{2}} dx \int_0^{\frac{1}{2}-x} \frac{(x-k)^2}{x\,(x+k)}\,dk +$$

$$+ \int_{\frac{1}{4}}^{\frac{1}{2}} dx \int_{\frac{1}{2}-x}^{x} \frac{(x-k)^2}{x\,(1-x-k)}\,dk + \int_{\frac{1}{2}}^{1} dx \int_0^{1-x} \frac{(1-x-k)^2}{(1-x)\,(1-x-k)}\,dk$$

(1) Infatti, dal punto di vista del Pietra, la massa dei casi possibili è espressa, nel 1° caso dall'integrale

$$\int_0^1 \int_0^{x(1-x)} dx\,dk = \int_0^1 x\,(1-x)\,dx = \frac{1}{6},$$

che rappresenta l'area racchiusa tra la curva $y = x\,(1-x)$ e l'asse della x, tra i punti 0,1 (v. nota 7), mentre dall'altro punto di vista, da me esposto nel testo, essa è espressa dall'integrale

$$\int_0^1 \int_0^{1-x} dx\,kx = \int_0^1 (1-x)\,dx = \frac{1}{2}$$

Nel 2° caso il Pietra calcola l'integrale:

$$\int_0^{\frac{1}{2}} \int_0^{x} dx\,dk + \int_{\frac{1}{2}}^{1} \int_0^{1-x} dx\,dk = \int_0^{\frac{1}{2}} x\,dx + \int_{\frac{1}{2}}^{1} (1-x)\,dx = \frac{1}{4}$$

che rappresenta l'area del triangolo di vertici (0,0), (1,0), $(\frac{1}{2}, \frac{1}{2})$, i cui lati tagliano, sulle parallele all'asse delle y, condotte per i punti $x = a$ segmenti uguali al capo di variabilità degli R o S corrispondenti a un $S_m = a$, mentre, dall'altro punto di vista, si calcolerebbe come nel 1° caso l'integrale

$$\int_0^1 \int_0^{1-x} dx\,dk = \frac{1}{2}$$

di concentrazione da lui stesso proposta in tale lavoro, gli indici R e S sono sempre concordi (1). Ma, esclusion fatta per la curva paretiana di 1ª approssimazione, che dipende da un parametro solo, per la quale l'asserto fu dimostrato dal Gini, in *Variabilità e Mutabilità*, la tesi è falsa. L'A. infatti deduce la concordanza dei due indici dal fatto che essi variano concordemente rispetto a *ciascuno* dei parametri dai quali dipendono. Ma ciò non significa nulla più di questo: *che se uno solo dei parametri varia, gli altri rimanendo costanti*, R e S crescono o diminuiscono entrambi. Che invece essi non possano essere sempre concordi quando, didipendendo da almeno due parametri, questi variino insieme, non è neppure necessario di dimostrarlo curva per curva, nei casi considerati dall'A., in quanto segue solo dal fatto che R e S siano due funzioni *diverse* di più parametri. Se, per esempio, questi sono in numero di due, per due punti corrispondenti generici R_0, S_0 delle superfici R e S esistono infinite direzioni, comprese tra le due, in generale diverse, che i piani orizzontali per i punti dati segnano sulle due superfici, spostandosi lungo le quali da R_0 e S_0, sull'una superficie ci si innalza e sull'altra ci si abbassa (2).

---

(1) Per le equazioni di Amoroso e di Gibrat cfr. L. Amoroso: *Ricerche intorno alla curva dei redditi*, già cit.; R. Gibrat: *Une loi des répartitions économiques: l'effet proportionnel*. « Bulletin de la Statistique Générale de la France », Tome XIX, Fasc. IV, Luglio-Settembre, 1930. L'equazione che l'A. propone per la rappresentazione delle curve di concentrazione dei redditi è $q = p - k\,p\,(1-p)^{\beta}$ ove p e q hanno il solito significato di frequenza rispettivamente dei possessori di un reddito inferiore ad un dato limite e del reddito complessivo corrispondente, e k e β sono due parametri variabili tra o e 1. Di essa non mi occupo perchè non rientra nei ristretti limiti di questa rassegna e per la stessa ragione non mi occupo del metodo dei semiinvarianti per la determinazione dei parametri della funzione di Mac Alister esposto dall'A. nello stesso lavoro.

(2) Le due direzioni tra le quali sono comprese quelle per le quali si verifica discordanza sono le direzioni per le quali si annullano i differenziali totali delle due funzioni $R(x\,y)$ e $S(x\,y)$: $\frac{d\,y}{d\,x} = -\frac{R'_x(x\,y)}{R'_y(x\,y)}$ e $\frac{d\,y}{d\,x} = -\frac{S'_x(x\,y)}{S'_y(x\,y)}$. Poichè $d\,R \gtrless 0$ a seconda che $\frac{d\,y}{d\,x} \gtrless -\frac{R'_x(x\,y)}{R'_y(x\,y)}$ e $d\,S \gtrless 0$ a seconda che $\frac{d\,y}{d\,x} \gtrless -\frac{S'_x(x\,y)}{S'_y(x\,y)}$, e poichè in generale $\frac{R'_x(x\,y)}{R'_y(x\,y)} \neq \frac{S'_x(x\,y)}{S'_y(x\,y)}$, per tutti i i valori $\frac{d\,y}{d\,y}$ compresi tra questi R cresce o diminuisce mentre S diminuisce o cresce. Se $\lambda(x\,y)$ è una funzione che assume in ogni punto valori compresi tra i due precedenti, ogni curva definita dall'equazione differenziale $y' = \gamma$

b) *Dipendenza di R e S dalla media aritmetica.*

Il D'Addario cerca di dimostrare che il rapporto di concentrazione « non prescinde dalla media aritmetica, anzi la sua definizione discende proprio dal concetto di *uniforme* distribuzione del carattere che si ha, evidentemente, quando ogni soggetto possiede un reddito esattamente uguale alla media aritmetica ». Dovremmo supporre che egli pensi che si possa costruire qualche indice di variabilità senza partire da quella « uniforme distribuzione » che deve essere in tale questione la pietra di paragone di ogni distribuzione. E che la quantità costante di carattere posseduta da ciascuna unità sia, nel caso di distribuzione uniforme,

---

è la proiezione sul piano $x\ y$ di due curve di discordanza corrispondenti, sulle due superfici. Tutte le coppie di punti $R\ S$, sufficientemente vicini, situate su tali curve, hanno come corrispondenti coppie discordanti di punti S o R. Per ogni punto delle due superfici passano infinite curve di discordanza corrispondenti. Nonostante il fatto che l'errore del D'Addario sia già stato brillantemente confutato anche dal Pietra (*A proposito della misura della variabilità e della concentrazione dei caratteri*, « Nuovi Problemi di Politica, Storia ed Economia » Anno VI, fascicolo 1-6, Ferrara, 1935), sia dal punto di vista teorico che dal punto di vista pratico, non mi è parso del tutto inutile riportare qui queste considerazioni che pur mi sembrano ovvie, in quanto la sicurezza con la quale l'A., affermando il suo dovere di esprimere e giustificare il suo pensiero nei riguardi della annosa disputa sulla scelta dei più convenienti indici atti a misurare la concentrazione dei redditi, dà per dimostrato ciò che non è vero, ha tratto in inganno più d'uno (cfr. ad esempio il lavoro di Dominedò, citato alla nota (1) e le recensioni del lavoro del D'Addario apparse in « Economia », Vol. XIV, n. 1, Luglio 1934, e nella « Rivista Bancaria », Anno XI, n. 1, Dicembre, 1934, di B. Barberi la prima e di G. Tagliacarne la seconda). La confusione è evidente. Il Tagliacarne scrive addirittura « . . . . . tanto il rapporto di concentrazione, rappresentato generalmente con S, quanto quello rappresentato con R sono la stessa cosa e conducono a risultati perfettamente concordanti . . . » ! È strano che il Santacroce, che pure ha recensito il lavoro (« Rivista Italiana di Statistica, Economia e Finanza », Anno VI, n. 5, Ottobre 1934) non mostri di accorgersi di un errore di tale entità. Segue però alla sua recensione una breve nota del D'Addario nella quale questi afferma che dal contesto del suo opuscolo si *arguisce* che egli intende considerare solo il caso in cui i singoli differenziali parziali presentino lo stesso segno [aumento (o diminuzione) generale della concentrazione]. Tale limitazione toglie però ogni valore alle conclusioni dell'A. Che essa si *arguisca* dal testo, a me non pare, nè si comprenderebbe, in tal caso, il motivo per il quale l'A. avrebbe ritenuta necessaria una simile precisazione.

uguale alla media aritmetica non può dirsi in generale più di quanto possa dirsi che è uguale alla mediana o ad un altro valore, medio od estremo, della distribuzione che si considera. Essa è la media aritmetica se si suppone costante l'ammontare totale del carattere ed il numero degli elementi tra i quali è ripartito. In tal caso, considerati unitari sia l'uno che l'altro, ogni distribuzione si pone a confronto con quella uniforme nella quale ogni frazione p del complesso (unitario) degli elementi della distribuzione possiede una uguale frazione q dell'ammontare totale (unitario) del carattere. Tale distribuzione di confronto è costante e quindi assolutamente indipendente da ogni caratteristica delle distribuzioni che ad essa si paragonano. Segue da ciò che tutti gli indici che si fondano sulla considerazione di una tale distribuzione uniforme facilmente si svincolano da ogni riferimento alla media aritmetica, se per mezzo di questa erano stati definiti (è il caso dell'indice S) e facilmente ad essa anche si riconducono quando erano stati definiti per via diversa (è il caso di R). La questione è ora superata nettamente, poichè, fin dal 1917, il Gini ha espresso lo scostamento semplice medio dalla media aritmetica come media dei valori assoluti degli scostamenti con ripetizione di un carattere nei singoli casi (1). D'altra parte, la differenza media fu già espressa, nella stessa memoria nella quale fu proposta, come media ponderata degli scostamenti assoluti dalla mediana (2), ed è noto come essa si possa esprimere come media degli scostamenti da una media qualsiasi (3). Ciò doveva rendere accorto l'A. sul significato da attribuire ad una qualsiasi espressione analitica del rapporto di concentrazione (4).

(1) Cfr.: C. Gini: *Di una estensione del concetto di scostamento medio e di alcune applicazioni alla misura dei caratteri qualitativi*, già cit., dove come scostamento con ripetizione di un carattere A nel caso i^{mo} è definita « la media aritmetica degli n scostamenti che il caso i^{mo} presenta rispetto a tutti gli n casi della serie » (pag. 405).

(2) Cfr. C. Gini: *Variabilità e Mutabilità*, già cit.

(3) Cfr., ad es., G. Mortara: *Lezioni di statistica metodologica* » dettate nel R. Istituto Superiore di Scienze Economiche e Commerciali di Roma, ediz. del « Giornale degli Economisti e Rivista di Statistica », Città di Castello, 1922, pag. 153.

(4) Non posso fare a meno di notare, data l'eco che le affermazioni dell'A. hanno trovato in varii recensori (Barberi, Tagliacarne, Santacroce), che la espressione alla quale ricorre l'A. nel tentativo di dimostrare il suo asserto :

$$R = \frac{2}{N A}\int_0^{\infty} \left\{ N A \int_0^{x} f(t)\, dt - N \int_0^{x} t f(t)\, dt \right\} f(x)\, dx$$

c) *Influenza del raggruppamento in classi della quantità di una distribuzione sulla determinazione di S e di R.*

È noto che, quando si conosca l'ammontare del carattere spettante complessivamente a ciascuna classe, influisce sulla determinazione di S solo l'errore che si commette nella scissione della classe, nella quale cade la media aritmetica, in due classi contigue aventi per estremo comune appunto la media aritmetica, mentre per la determinazione di R occorre conoscere la distribuzione del carattere entro ciascuna classe. Non è dubbio perciò che R debba risultare affetto in generale, per questa circostanza, da un errore maggiore che non S (1). Ma al problema teorico non accenna affatto l'Yntema, che è il solo dei nostri Autori che si occupi dell'argomento. Egli mette a confronto i valori che gli otto indici di variabilità da lui considerati assumono per una distribuzione di 100 redditi individuali quando si conosce l'altezza di ciascun reddito e quando si conosce solo la distribuzione per classi di reddito, dei redditieri e del reddito loro spettante. Per il rapporto di concentrazione il valore relativo alla seriazione completa supera quello relativo all'altra, e la differenza (0.014) occupa il 4° posto nella graduatoria crescente delle corrispondenti differenze di tutti gli 8 indici. A parte il fatto che un esempio solo, in statistica, non è significativo, nulla sarebbe da obbiettare se l'Yntema non avesse calcolato il rapporto di concentrazione con una formula evidentemente er-

---

nulla dice più di quanto dica la espressione più semplice $R = \int_0^1 (p - q)\, dp$, dalla quale parte, per pervenire a quella che egli giudica si presti meglio ai suoi ragionamenti. Entrambe le funzioni integrande rappresentano infatti « la deviazione fra ciò che dovrebbero possedere i possessori di un reddito non maggiore di x, nel caso di equiripartizione, e l'ammontare complessivo dei redditi da essi effettivamente posseduti ». Ma se l'autore sa *per precedenti cognizioni*, « ciò che dovrebbero possedere i possessori di un reddito non maggiore di x nel caso di equiripartizione » perchè si pone subito dopo questa domanda e trae dalla sua espressione la risposta che aveva dovuto ammettere di sapere per aver ragione di porsi la domanda ? Era, del resto, naturale che una dimostrazione che doveva logicamente ridursi alla sola osservazione — che abbiamo già notato come non significhi nulla — che anche R può esprimersi come somma di scostamenti dalla media aritmetica, dovesse ridursi ad un giro di parole quanto mai oscuro e faticoso appunto perchè privo di significato.

(1) Cfr. C. Gini : *Variabilità e Mutabilità*, già cit., pagg. 96-97.

rata per difetto quando le distribuzioni sono date per classi, e tanto più errata quanto minore è il numero delle classi : il difetto in tal caso non è imputabile all'indice, ma alla formula scelta per calcolarlo. La formula è $R' = \frac{1}{2} \Sigma (p_{k-1} q_k - p_k q_{k-1})$ avendo p_k e q_k l'ordinario significato di frequenza, rispettivamente, dei redditieri con un reddito inferiore ad un dato limite e dell'ammontare complessivo dei redditi corrispondenti. Essa corrisponde all'ipotesi che i redditi di ciascuna classe siano uguali, cioè, che essi siano distribuiti, nel complesso, meno disugualmente che in realtà. Il Gini, che si era posto già in *Variabilità e Mutabilità* il problema della determinazione degli indici di variabilità quando le quantità della seriazione siano raggruppate in classi, ha studiato poi per esteso il problema (1) per quanto riguarda il rapporto di concentrazione, giungendo a formule che hanno fornito approssimazioni soddisfacentissime. Anche la determinazione per via puramente grafica di R, fondata sulla misura planimetrica dell'area di concentrazione, può fornire valori ottimamente approssimati di R (2). L'Yntema non sembra conoscere tali ricerche (3).

---

(1) Cfr. C. Gini: *Sulla misura della concentrazione e della variabilità dei caratteri* », già cit.

(2) Cfr. C. Gini: *Intorno alle curve di concentrazione.* « XX Sessione dell'Istituto Internazionale di statistica », Madrid, 1931, e ripubblicata in « Metron », Vol. IX, n. 3-4, 1932.

(3) La osservazione del testo non sminuisce però uno dei maggiori pregi del suo lavoro, che è quello di interessare il pubblico degli statistici anglo-sassoni ad una questione d'indubbia importanza quale è quella della scelta dell'indice di variabilità, così discussa da noi. Ed è non senza un intimo compiacimento che si nota la buona conoscenza, anche se incompleta, che l'A. dimostra delle principali fonti nostre originali da lui citate; buona conoscenza che, se è un dovere dello studioso, non sembra certo essere pregio costante degli Autori nord-americani. Capita, infatti, di trovare nello stesso numero del Giornale, un articolo di Miss. Barbara Burks, fondato su un metodo già, proposto ed applicato dal Gini nel 1910 (e forse anche prima ancora, da altri), e nel numero di giugno dello stesso anno, un articolo di G. P. Watkins nel quale è proposto un indice ordinale di correlazione che altro non sembra che una edizione riveduta e scorretta dell'indice di cograduazione, dallo stesso Gini proposto sino dal 1914, e largamente usato in Italia. In nessuno dei due articoli è neppure citato il nostro Autore. (Cfr. Miss. Barbara Burks: *A statistical method for estimating the distribution of sizes of completed fraternities in a population represented by a random sampling of individual.* « Journal of American Statistical Associa-

Sulla misura più opportuna della variabilità.

Questa questione ebbe origine, si può dire, dalla introduzione della differenza media nella metodologia statistica.

Al suo primo apparire, in *Variabilità e Mutabilità*, nel 1912, la differenza media incontrò subito largo favore tra gli statistici, e esposero il procedimento per determinarla, o l'applicarono moltissimi autori quali: Savorgnan, Dettori, Porru, Camboni, Pietra, Mortara, Niceforo, De Pietri Tonelli in Italia (1) e Lexis, Czuber, Weinberg, Dalton, Bowley, Julin, March all'estero (2), per non citare che i più antichi, e non tutti.

---

tion », n. 184, dicembre 1933; G. P. Watkins: *An ordinal index of correlation.* « Ibidem », n. 182, Giugno 1933; e la lettera del Gini pubblicata nel numero di Giugno 1934 dello stesso Giornale.

(1) Cfr. F. Savorgnan: *La distribuzione dei redditi nelle provincie e nelle grandi città dell'Austria*, già cit.; *La ripartizione generale della ricchezza privata italiana*, « Metron », Vol. I, n. 3, 1921; G. Dettori: *Contributo allo studio della variabilità dei prezzi.* « Studi economico-giuridici della R. Università di Cagliari », Parte I, Vol. IV, 1912, e *Lezioni di statistica economica*, nel R. Istituto Superiore di studi economici e commerciali, Genova, 1917-18; E. Porru: *La concentrazione della ricchezza nelle diverse regioni d'Italia*, già cit.; L. Camboni: *Della correlazione tra alcuni fenomeni economici e sociali e la criminalità.* Cagliari, Società Tipografica Sarda, 1913; F. P. Cantelli: *Sulla differenza media con ripetizione.* « Giornale degli Economisti e Rivista di Statistica », Febbraio, 1913; G. Pietra: *Appunti intorno alla misura della variabilità e alla concentrazione dei caratteri*, Roma, Bertero, 1915 e *The theory of statistical relations with special reference to cyclical series.* « Metron », Vol. V, n. 3-4, 1924; G. Mortara: *Elementi di statistica*, « Athenaeum », Roma, 1917 e *Lezioni di statistica metodologica*, Città di Castello, 1922; A. Niceforo: *La misura della vita.* Bocca, Torino, 1919; A. De Pietri Tonelli: *Il commercio estero sotto l'aspetto statistico, Lezioni di Politica Commerciale.* « R. Scuola Superiore di Commercio », Venezia, 1919-20 e *Lezioni di Scienza economica razionale e sperimentale.* Industrie Grafiche Italiane, Rovigo, 1920; *I tassi ufficiali di sconto nelle principali piazze europee negli ultimi dieci anni*, « Rivista di Politica Economica », 1923; *La lira italiana*, ibidem, 1925.

(2) Cfr. W. Lexis: *Jahrbücher fur Nationalökonomie und Statistik*, September 1914; E. Czuber: *Beitrag zur Theorie der statistischer Reihen*, Wien, 1914; W. Weinberg: *Ueber Korrelationsmessung.* « Deutsches Statistisches Zentralblatt », Juni-Juli 1916; H. Dalton: *The measurement of the inequality of incomes*, « The Economic Journal », September, 1920; A. L. Bowley: *Elements of statistics*, London, King, 1920; A. Julin: *Principes de statistique théorique et appliquée*, Tome Ier, Statistique théorique, Paris, Rivière, 1921; L. March: *Les principes de la méthode statistique*, Paris, 1930.

Il Gini aveva impostato chiaramente in *Variabilità e Mutabilità* il problema della scelta dell'indice di variabilità come un problema la cui soluzione dipenda essenzialmente dallo scopo della ricerca. Furono altri Autori subito dopo di lui, quali lo Czuber, il Weinberg, il Pietra, e più tardi il Mortara, a ritenere invece che il fatto che la differenza media elimini, nei confronti degli « scostamenti medi » l'arbitrio della scelta della media da cui i singoli scostamenti sono misurati, potesse essere una ragione per preferire in generale la differenza media agli altri indici di variabilità fondati sugli scostamenti da una media (1) ; e analogamente il Pietra aveva ritenuto che potesse costituire ragione di preferenza per la differenza media la circostanza che il campo di variazione di questa è maggiore, a parità di scostamento semplice medio dalla media aritmetica o dalla mediana, del campo di variazione dello scostamento semplice medio dalla media aritmetica o dalla mediana, a parità di differenza media (2). Ragioni di preferenza queste indubbiamente giuste « *coeteris paribus* », e cioè o per la mancanza di ragioni specifiche che potessero consigliare, secondo il concetto del Gini, un indice piuttosto che un altro, o per la perfetta equivalenza di tutti gli indici — casi questi di nessun interesse il primo, perchè presuppone una indagine così grossolana sulla variabilità che tutti gli aspetti di questa risultino equivalenti ai fini dell'A., d'interesse esclusivamente teorico il secondo, perchè presuppone una legge analitica *precisa*, esprimibile mediante un solo parametro — ma non giuste « in generale », e certamente suscettibili di errata interpretazione, in quanto possono ingenerare il convincimento che gli indici di variabilità siano da considerarsi piuttosto come strumenti più o meno perfezionati per la necessità di una stessa opera, che come strumenti diversi, più o meno simili, adatto ciaciascuno alla necessità di un'opera diversa.

Proprio un tale convincimento sembrava avere il Ricci quando affermava che è tempo male impiegato andare in traccia di altri

---

(1) Cfr. E. Czuber : *Beitrag zur Theorie*, ecc., già cit. ; W. Weinberg : *Ueber Korrelationsmessung*, già cit. ; G. Pietra : *Recenti pubblicazioni di metodologia statistica*, in « Rivista Italiana di Sociologia » Marzo-Giugno, 1917; G. Mortara : *Lezioni di statistica metodologica*, già cit.

(2) Cfr. G. Pietra : *Delle relazioni tra gli indici di variabilità*, già cit. ; *Sulla variabilità delle serie statistiche*. « Rivista Italiana di Sociologia », Maggio-Agosto 1915.

indici di variabilità dal momento che già uno ne esiste (^1S_A), ed è ottimo (1).

E dallo affermare ciò a capovolgere la tesi Czuber-Weinberg-Pietra che la indipendenza dalla media aritmetica costituisca una ragione di preferenza per la differenza media è un brevissimo passo che il Ricci compie, sull'esempio del Bresciani (2), al quale sembra più giusto assumere, come misura della disuguaglianza, la somma degli scostamenti dallo stato di uguaglianza definito dalla media aritmetica, che non la somma delle differenze tra tutte le quantità. E a confortare la legittimità dello scostamento semplice medio, da tale punto di vista, entrambi gli autori si rifanno alle opinioni espresse dal Mortara nel 1911 (3), mentre, a sua volta, il Mortara, sebbene abbia riconosciuto in seguito la differenza media come la misura più corretta della disuguaglianza (4), fa oggi sue (5) le vecchie idee del Bresciani. Su questo terreno lo segue D'Addario (6), che sostenendo, come abbiamo visto, che il rapporto di concentrazione discende proprio dal concetto di uniforme distribuzione, pensa di apportare il colpo di grazia alla differenza media, distruggendo « l'unico preteso vantaggio di R su S ». Sorge il dubbio che, per reazione al suo antico amore per la differenza media (7), sia il D'Addario tanto cattivo con questa e così aggressivo con quanti hanno il torto di gradirla. Curioso è vedere

---

(1) Cfr. U. Ricci : *L'indice di variabilità e la curva dei redditi.*«Athenaeum », Roma 1916, e cfr. anche la replica del Pietra, citata più avanti (pag. 159).

(2) Cfr. C. Bresciani-Turroni : *Recenti pubblicazioni.* « Giornale degli Economisti e Rivista di Statistica », Gennaio, 1916.

(3) Cfr. G. Mortara : *Note di economia induttiva,* già cit.

(4) Cfr. G. Mortara : *Lezioni di statistica metologica,* già cit.

(5) Cfr. G. Mortara : *Sulle disuguaglianze statistiche,* già cit.

(6) Cfr. R. D'Addario : *Sulla misura della concentrazione dei redditi,* già cit.

(7) Scriveva infatti una volta il D'Addario: « Tra i molti indici sintetici possibili noi preferiamo il rapporto di concentrazione del Gini e la metà della deviazione media relativa dalla media aritmetica . . . . Essi sono da preferirsi perchè prescindono dalla forma della distribuzione e perchè hanno un significato chiaro, semplice e preciso». (Cfr. *Un'indagine sulla dinamica distributiva,* già cit., pagg. 282-285)». L'indice S, però, ci dà un'idea troppo sommaria della distribuzione. Valori uguali di S. possono evidentemente risultare da distribuzioni fra loro molto dissimili ». (Cfr. : *La distribuzione dei redditi mobiliari in terra di Bari.* « Archivio Scientifico », R. Istituto Superiore di Scienze Economiche e Commerciali di Bari, Vol. II, 1928-29, pag. 349.

quanto impegno egli ponga sia nel difendere la Modestia, offesa da qualche Autore (1), sia nello « affrontare e chiarire questioni sulle quali più acuto era e rimane il dissenso con alcuni Autori ». Senonchè sembra una affrettata risposta al Pietra, la tesi del D'Addario che almeno per quanto concerne le equazioni sinora proposte per rappresentare analiticamente le distribuzioni dei redditi, gli indici S e R conducono alle stesse conclusioni. Sembrerebbe anche questa di una tale concordanza tra R e S, un'idea gettata giù solo per fare un dispetto al Pietra, se essa non finisse col fare un più grosso torto al suo Autore.

Ma poichè non si può, come abbiamo visto, assolutamente parlare di concordanza tra R e S, resta escluso che l'uso dell'uno o dell'altro sia indifferente (2) e che il problema della scelta si ri-

---

(1) Il D'Addario dichiara (*Sulla concentrazione dei redditi*, già cit., pag. 26) che lo infastidisce la pretesa del Pietra, che si attribuisca a lui, che l'ha dimostrata, la relazione tra lo scostamento semplice medio dalla media aritmetica di una distribuzione, ragguagliato al suo massimo e la massima differenza tra i valori di p e di q corrispondenti. Pretendere ciò è pretendere il proprio, e non pretendere la parternità dell'indice S, come afferma il D'Addario. Il rilievo del Pietra (*Intorno alle doscordanze fra gli indici di variabilità e di concentrazione*, già cit., nota (2), pag. 5-6 dell'estratto) all'Yntema, che sembra attribuire tale relazione al Ricci, più che inopportuno sembra giustificato, in quanto lo stesso errore dell'Yntema era già stato, commesso dal v. Bortkiewicz (*Die Disparitätsmasse in der Einkommensstatistik*, già cit.), come fu rilevato dal Gini (*Observations à la communication: Die Disparitätsmasse in der Einkommensstatistik del Prof. L. von Bortkiewicz*, già cit.). A parte ciò, non si capisce la ragione del trattamento specialissimo che il D'Addario fa al Pietra quando afferma che, neanche se questi avesse proposto lui lo scostamento semplice medio dalla media aritmetica, gli si potrebbe riconoscere la paternità di questo indice, ma sarebbe anche necessario per questo che egli avesse insegnato « come si calcola un qualsiasi coefficiente di disuguaglianza » e gli si farebbe ancora un favore !

(2) Non comprendo come si accordi questa equivalenza di R e S (ma equivalenti da che punto di vista ?) con la distinzione tra R e S *misure* e α *indice* di concentrazione che l'A. fa a pag. 32, a proposito della curva paretiana di prima approssimazione. Se le due *misure diverse* R e S sono equivalenti per l'A. ciò vuol dire che egli non dà peso alla intensità delle variazioni di R e S, ma solo al segno di tali variazioni. Egli cioè, non usa R e S per *misurare* la concentrazione, ma solo per giudicare della maggiore o minore concentrazione di due diverse distribuzioni; uso al quale, nel suo caso, basta benissimo il molto più semplice indice α o qualsiasi funzione monotona di α, indipendentemente dal suo significato. Che se poi concepisse R e S come vere misure, cioè come la precisa espressione quantitativa della concentrazione, l'uso dell'uno o dell'altro non sarebbe più affatto indifferente.

duca ad una questione « di convenienza e di opportunità » come l'A. dichiara. Quali siano queste ragioni « di convenienza e di opportunità » a favore dell'uno o dell'altro indice, che noi potremmo accettare almeno per la curva paretiana di prima approssimazione, non appare molto chiaro dal testo (1) e contro di esse sarebbero in ogni caso quelle considerazioni di indole generale, che l'A. pensa dovrebbero indurre senz'altro a preferire, come egli preferisce, S a R. Queste considerazioni « di indole generale » sono in tutto due, e cioè che S è « più semplice nell'applicazione » e « meglio corrispondente alle consuetudini della pratica e della scienza ». Delle due, la prima costituisce un criterio che può valere solo *coeteris paribus* (2), ciò che finora non può dirsi che sia e al quale, del resto, la crescente diffusione e perfezione dei mezzi meccanici di calcolo, e la possibilità di determinare per via grafica R, tolgono ogni importanza; la seconda è una frase del Mortara e ci richiama alle idee svolte da quest'ultimo nella memoria *Sulle disuguaglianze statistiche* della quale discorreremo poco più oltre.

Più razionale sembra, al confronto, l'idea di Yntema (3) che

---

(1) L'A. in un lavoro *Intorno alla curva dei redditi di Amoroso* pubblicato in « Rivista Italiana di Statistica, Economia e Finanza », Anno IV, n. 1, Marzo 1932, nel quale commette lo stesso errore di ritenere perfettamente analogo il comportamento di R e S in un caso particolare della curva di Amoroso, nel quale R e S possono considerarsi funzioni di due parametri, dichiara che la scelta deve essere determinata « di volta in volta da ragione di opportunità e di *facilità di calcolo* ». Sembra quindi che le « ragioni di convenienza siano tutt'uno con quella ragione di indole generale » di maggiore semplicità nell'applicazione, della quale l'A. parla nella Introduzione. E sembra strano che attribuisca tanto peso ad una considerazione del genere egli che invita (pag. 18) gli « statistici di buona volontà e di buon coraggio » ad usare per la rappresentazione analitica della distribuzione dei redditi, la curva dell'Amoroso che se anche non si vuol giudicare, col Pietra, « praticamente inutilizzabile dagli statistici », non si può negare certo che sia, rispetto ad altre, di tanto più laboriosa applicazione.

(2) Anche lo scostamento quadratico medio esige calcoli più laboriosi che non lo scostamento semplice medio, e non per questo certo è stato bandito dall'uso. E il calcolo della differenza media non è certo tanto più faticoso di quello dello scostamento quadratico medio (v. al riguardo l'opinione espressa dall'Anderson nel suo recentissimo libro: « *Einführung in die Mathematische Statistik*, già cit.) e più di un metodo è stato proposto per facilitarlo (cfr. De Finetti : *Sui metodi proposti per il calcolo della differenza media.* « Metron », Vol. IX, n. 1, Roma, 1932).

(3) Cfr. D. B. Yntema : *Measures of the inequality in the personal distribution of wealth or income*, già cit.

possa costituire ragione di preferenza per un indice la sua minore sensibilità alla influenza perturbatrice esercitata dalla circostanza che, nella generalità dei casi, non si conoscono le intensità del carattere di ciascuna unità statistica, ma si conosce solo la distribuzione di tali unità e dell'ammontare totale del carattere, entro classi, più o meno ampie, della intensità del carattere stesso.

Ma neppure questa idea tocca il vivo della questione, nè era intenzione dello Yntema, mi pare, di risolvere alcunchè, bensì solo dimostrare la confusione che deriva in pratica dall'uso di molteplici indici.

Ben otto indici diversi egli ha usato per misurare la disuguaglianza di alcune distribuzioni della ricchezza e del reddito, mostrando come tra le graduatorie dei valori assunti da tali indici sia per le distribuzioni della ricchezza, sia per le distribuzioni del reddito, concordino ottimamente quella dei valori dello scostamento semplice medio ragguagliato alla media aritmetica con quella dei valori del rapporto di concentrazione ; quelle dei valori dello scostamento quadratico medio ragguagliati una volta alla media aritmetica ed un'altra alla media quadratica delle distribuzioni, tra di loro, e similmente, tra di loro, quelle dei valori degli scostamenti semplice e quadratico medio non delle distribuzioni delle intensità dei due caratteri, ma dei logaritmi di tali intensità. Le graduatorie di ciascuno di questi tre gruppi discordano sensibilmente da quelle degli altri due e dalle graduatorie dei valori corrispondenti degli indici α, del Pareto, e δ, del Gini, le quali non concordano nè tra di loro nè con nessun'altra.

Si potrebbe obbiettare, limitatamente alla esemplificazione addotta dall'Yntema, che gli indici α essendo subordinati alla possibilità di rappresentare più o meno bene, con le corrispondenti curve, le distribuzioni empiriche cui si riferiscono, sembra escluso che si possa attribuire loro, nel suo caso, un significato molto preciso (1). E si potrebbe anche osservare, in generale, che non si può, a buon diritto, ascrivere la lamentata confusione alla molti-

---

(1) Sebbene non possegga i dati interpolati dall'A., ritengo l'affermazione del testo sufficientemente autorizzata dal confronto, da me eseguito, tra i valori S, R, δ e S, R, α da esso calcolati per ciascuna distribuzione e i corrispondenti che si ricavano rispettivamente dai valori di α e di δ nell'ipotesi che le relative curve di ripartizione e di concentrazione rappresentassero esattamente i dati empirici. Il risultato del confronto non è

tudine degli indici cui fanno ricorso i vari Autori, ma piuttosto questa moltitudine di indici alla confusione dei vari Autori i quali, spesso, non sanno con sufficiente precisione che cosa vogliano misurare. Ma avrò agio di tornare su questo argomento, a maggior proposito, poco più oltre.

Dove Yntema ha però, indubbiamente, sfiorato il centro della questione è stato quando ha scritto che la scelta tra i vari indici « is largely determined by . . . . the tacit assumption that inequality consists in that aspect of personal distribution which a given coefficient measures ». È appunto in questa identificazione arbitraria del concetto generale con le sue determinazioni particolari che sta la ragione della non più pacifica convivenza della vasta famiglia degli indici di variabilità, chiamati a sostenere ciascuno un ruolo più grande di sé, e quindi contrapposti l'un l'altro come una cosa alla cosa stessa. Non era sfuggito al Gini, anzi egli aveva esplicitamente avvertito che è lo scopo della ricerca che deve suggerire l'indice da adottare (1), ed egli aveva anche ammonito che lo statistico « would be seriously mistaken if he pretended to constrain the multiform reality to lie on the Procustean bed of one or a few indices (2) ; e della stessa idea era il Pietra, quando spiegava al Ricci che se « per via di indici diversi, ciascuno dei quali venga a caratterizzare di un fenomeno statistico una diversa proprietà, possiamo guardare il fenomeno stesso sotto diversi punti di vista, non avremo a dolerci che lo statistico impieghi il suo tempo a fabbricare indici, ma anzi a congratularcene, in quanto che egli avrà resi così i metodi della statistica metologica rispondenti alle questioni pratiche che di volta in volta esigono di considerare nei fenomeni l'uno o l'altro aspetto » (3). È certo che a chi non ha chiaro

---

confortante. A bene intendere, ciò non significa affatto che la curva di Pareto, o quella del Gini, non possano descrivere, anche bene, le distribuzioni osservate, ma significa solo che le curve interpolate dall'A. non riescono a ciò. Poichè non è escluso che un metodo d'interpolazione diverso conduca a valori di α e δ più soddisfacenti.

(1) Cfr. C. Gini : *Variabilità e mutabilità*, già cit.

(2) Cfr. C. Gini : *Contributions of Italy to modern statistical methods*, « Journal of the Royal Statistical Society ». Vol. LXXXIX, Part IV, July, 1926.

(3) Cfr. G. Pietra : *Recenti pubblicazioni di metodologia statistica*. « Rivista Italiana di Sociologia », Marzo-Giugno 1917, v. pure : *Note di statistica metodologica : A proposito della misura della variabilità e della concentrazione dei caratteri*, già cit., nel quale è ribadito lo stesso concetto (pag. 6 dell'estratto).

il concetto della diversità sostanziale dei vari indici, deve apparire logico di contrapporli l'un l'altro quando si notino tra essi delle discordanze, e di considerare ogni avvertimento in contrario o come giustificazioni generiche di indici nuovi o come semplici, seppur brillanti, battute polemiche.

Anche il Mortara ha sfiorato il centro della questione, se nella già citata memoria *Sulle disuguaglianze statistiche* ha tentato di di risolvere il problema risalendo ai primi principi. Ma la soluzione data dal Mortara è essenzialmente dommatica.

Egli dà due definizioni della « disuguaglianza fra più dati numerici : *A*) la deviazione di ciascuno dei dati in esame da un livello ipotetico di uguaglianza, che normalmente si fa corrispondere ad una media dei dati stessi (per lo più alla media aritmetica) ; *B*) la deviazione di ciascuno dei dati in esame da ciascuno degli altri. Sia col criterio *A*, che è un criterio di comparazione indiretta, sia col criterio *B*, che è un criterio di comparazione diretta, si determinano tante disuguaglianze singole, la somma dei cui valori assoluti fornisce la disuguaglianza totale : questa si divide per il numero delle singole disuguaglianze per svincolarla dal numero dei dati in questione ; si ragguaglia poi ad una media per misurarne l'importanza, oppure si ragguaglia al massimo valore che si sarebbe potuto presentare in una ipotesi determinata — della costanza, generalmente, del numero e della somma dei dati — per misurarne il grado. Tutte le misure della disuguaglianza discendono da quelle due definizioni. Quando i dati da comparare sono soltanto due, più semplice nell'applicazione e più intuitivo nel risultato si presenta il criterio *B* ; quando i dati sono più di due il criterio *A*. Segue che la misura adatta delle disuguaglianze statistiche è lo scostamento semplice medio.

È qui « the tacit assumption » dell'Yntema diventata duplice categorica esplicita affermazione ; è la difesa questa delle vecchie idee del Bresciani, è la vecchia via dell'unico indice battuta dal Ricci, che ripercorre, non importa con quali mezzi, l'Autore.

E sommamente speciosa appare l'argomentazione che permette al Mortara di eliminare in pratica il dualismo dei due principi dai quali parte in teoria.

Sembra per lo meno dubbio, per es., che, in una popolazione di 1000 individui, la quale disponga di un reddito complessivo di 10.000.000, del quale la metà appartenente ad un individuo solo, e l'altra metà ugualmente divisa tra gli altri 999 individui, chi

voglia rendersi conto della propria situazione economica, pensi di paragonare le proprie 5.005 lire di reddito alle astratte 10.000, che non possiede nessuno, anzichè alle stesse 5.005 lire di 998 dei suoi concittadini ed ai 5 milioni dell'altro (1). E sembra anche dubbio di poter trovare argomenti abbastanza solidi per convincere di errore un individuo che tra la popolazione considerata ed un'altra ugualmente numerosa e di uguale reddito complessivo, una metà del quale appartenente ad un individuo solo e l'altra metà ugualmente divisa tra 500 individui, 499 individui non possedendo reddito alcuno, voglia giudicare la disparità della distribuzione maggiore nella seconda.

Ma non è che l'A. si accinga ad un'impresa del genere. Egli afferma solo che, in base ad un criterio *A* quell'individuo ha torto perchè la disuguaglianza delle due distribuzioni è la stessa ; in base ad un criterio *B* quell'individuo ha ragione perchè la disuguaglianza della seconda distribuzione è maggiore di quella della prima ; ma che è più conveniente dire che egli ha torto, perchè quel valore comune che il criterio *A* assegna alle due disuguaglianze si calcola più facilmente (2).

Appare giustificata qualche riluttanza ad accettare un principio del genere.

E appare tanto più giustificata in quanto anche gli Autori che formalmente l'accettano in teoria non se ne tengono affatto paghi in pratica, quali il D'Addario e il Dominedò.

Curiosa è la posizione concettuale di quest'ultimo Autore (3), che parte dai concetti esposti dal Mortara, accettando, almeno in linea di principio, la legittimità dell'uso dello scostamento semplice medio (che non credo contestata in linea di principio da nessuno, tanto meno dal Gini, il quale distingue e propone l'uso cosciente dei vari indici a seconda dei casi e a seconda degli scopi della ricerca : cfr. *Variabilità e Mutabilità*, già cit.) in con-

(1) Il March afferma, molto più recisamente, la stessa cosa : « L'individu ne peut guère que comparer son salaire à celui de ces compagnons. On est, dès lors, conduit à considérer les différences de chaque salaire avec tous les autres, les différences étant formées de toutes les manières possibles et prises en valeurs absolues ».

(L. March : *Les principes de la méthode statistique*. già cit., pag. 295).

(2) L'altra ragione con la quale l'A. giustifica il criterio *A*, nel caso di più dati, e cioè quella « di essere più intuitivo nel risultato », non avrebbe evidentemente valore per l'individuo che ho supposto nel testo.

(3) Cfr. V. Dominedò : *Sulla differenza media*, già cit.

trapposto ad un meno saldo fondamento logico della differenza media, per poi concludere affermando la opportunità pratica di « conservare » quest'ultima quale utile strumento di indagine atto a supplire ad un difetto che sarebbe insito nello scostamento semplice medio « se calcolato una volta soltanto per l'intera serie ». Criticando la differenza media perchè « a chiunque rifletta sul modo di calcolarla, appare in contrasto con un ovvio criterio di simmetria, giacchè sintetizza confronti istituiti fra ciascuno dei dati in esame ad un complesso di termini il quale normalmente *non* è uguale per tutti » (1), sembra che debba coerentemente rigettare l'applicazione della definizione B di disuguaglianza che rappresenta un criterio di comparazione diretto dei dati, il quale giustifica, secondo il Mortara, la differenza media, rigettarla, dico, anche nel caso di due soli dati statistici, nel quale esso « si presenta come il più semplice nell'applicazione e il più intuitivo nel risultato » giusto dice il Mortara. Non solo egli non respinge il criterio B, ma nella conclusione parla invece di un difetto insito in S, mentre sarebbe forse più giusto affermare che, data la definizione dalla quale discende, S non tiene conto di particolarità delle distribuzioni che, essendo essenziali in talune ricerche, consigliano senz'altro l'adozione di Δ.

Data la posizione concettuale dell'A. che sembra accettare la impostazione del problema fatta dal Mortara e poi critica, già in linea di principio, la definizione di disuguaglianza dalla quale questi deriva la differenza media, era logico che questo A. trovasse « qualche cosa d'ingiustificato se non addirittura di arbitrario » nella formazione di Δ (2). Più interessante è la « giustificazione in-

---

(1) Che la maniera di calcolare la differenza media sia in contrasto con un « ovvio criterio di simmetria », io nego recisamente. O che forse non è perfettamente simmetrico il procedimento quando i termini essendo due soli, si paragoni A a B e B ad A ? Nel caso della differenza media con ripetizione poi, le obbiezioni, già infondate per la differenza media, dileguano del tutto. Nei riguardi del rapporto di concentrazione, essa è ugualmente insussistente potendosi questo riguardare come la differenza media con ripetizione ragguagliata al suo massimo.

(2) La formula del Dominedò, che esprime la somma dei valori assoluti degli scostamenti del caso iesimo da tutti gli altri quando l'intensità del carattere sia, nel caso iesimo, superiore alla media aritmetica, si ricava senza fatica dalla (25) di *Variabilità e Mutabilità* (pag. 36). E anche l'esame del « perchè e di quanto » differiscono tra loro S e Δ_R appare già in *Variabilità e Mutabilità* attraverso il confronto di S e di Δ_R con lo sco-

diretta » che egli dà di questa. Tale giustificazione si fonda sul fatto che la differenza media si può esprimere come somma, opportunamente ponderata, degli scostamenti di 1°, 2°, ... , n° ordine, relativi, rispettivamente, alla serie che si considera ; alle due serie parziali nelle quali questa è bipartita dalla media aritmetica ; e così via via alle serie di ordine n nelle quali risultano bipartite dalla propria media aritmetica le serie di ordine $n - 1$.

Questa considerazione discende immediatamente dalla interpretazione geometrica della differenza media e dello scostamento semplice medio dovute la prima al Gini (1), la seconda al Pietra (2), e risulta dalla scomposizione dell'area di concentrazione nei triangoli di massima altezza aventi il vertice sulla curva di concentrazione e per base la retta di equidistribuzione il primo, i lati di questo i due successivi, e così via, già considerata, per altri scopi, dal Gini (3). Avere messo in luce il significato di tale scomposizione dal punto di vista della variabilità costituisce, ad ogni modo, un contributo positivo alla questione.

Ma la impostazione del problema, fatta dal Mortara, solleva anche un'altra questione. Solleva la questione se sia legittimo definire una stessa cosa in due maniere diverse. È vero che nella scienza si incontrano sovente definizioni varie di una stessa cosa, ma si tratta sempre di definizioni solo formalmente diverse, mettendo ciascuna particolarmente in luce una proprietà caratteristica dell'oggetto definito dalla quale seguono tutte le altre. Le varie definizioni sono concettualmente identiche e praticamente equivalenti. La ragione della preferenza che si accorda all'una o all'altra di esse, a seconda dei casi, risiede in un principio di economia del pensiero, il quale attua, nel suo campo, la legge del minimo sforzo, scegliendo, tra tante definizioni, che gli sono, in linea di principio, indifferenti, quella che più direttamente porta al suo scopo. Quando tale equivalenza non esiste, come per le varie de-

---

stamento semplice medio dalla mediana (pagg. 33-35). L'A. sembra non conoscere affatto *Variabilità e Mutabilità* e ciò potrebbe essere effetto della rarità dell'opera, già denunciata dal Bortkiewicz. Tanto più mi pare oggi che un'altra edizione di quest'opera, nel volume, sotto stampa, nel quale sono raccolte le principali memorie di metodologia statistica del Gini, risponda ad una sentita necessità.

(1) Cfr. C. Gini : *Sulla misura della concentrazione e della variabilità dei caratteri*, già cit.

(2) Cfr. G. Pietra : *Delle relazioni tra gli indici di variabilità*, già cit.

(3) Cfr. C. Gini : *Intorno alle curve di concentrazione*, già cit.

finizioni della media aritmetica nel caso di caratteri qualitativi (1), esse non possono essere accettate insieme.

Ed è proprio questo il caso del Mortara, perchè le due definizioni della disuguaglianza, sebbene concettualmente meno diverse di quel che appare a prima vista, essendo possibile svincolare, dal riferimento alla media aritmetica quella che discende dal criterio *A* e vincolare a tal riferimento quella che discende dal criterio *B* (2), pure non sono affatto equivalenti, in quanto tengono conto diversamente degli stessi elementi costitutivi. Tra di esse esiste quindi vera opposizione e si prova una certa difficoltà ad intendere perchè una mente scientifica dovrebbe ammetterle entrambe, riuscendo spesso a risultati contradittori, e dovrebbe poi scegliere tra di esse, in base ad esigenze puramente formali, quale potrebbe essere quella di una maggiore speditezza di calcolo. Ed è assurdo che la definizione valida nel caso di due dati non serva più quando i dati siano più di due (3).
Nè si vede perchè si dovrebbe rinunziare, anche in linea di principio, ad una piuttosto che all'altra delle due definizioni. La contraddizione non pare evitabile.

E non è evitabile finchè, coesistendo nell'uso molteplici misure, ciascuna delle quali presenta vantaggi suoi particolari, si voglia sostenere che tutte rappresentano la stessa cosa. Non sarebbe, invece, errato affermare — come ha ribadito recentemente anche il Pietra (4) — che rappresentano ciascuna un aspetto diverso dalla stessa cosa. Questa potrebbe essere la *variabilità*, cioè l'attitudine di un carattere ad assumere appunto modalità quantitative (5). Così definita la variabilità è un concetto, non una quantità, suscettibile di *misura*.

---

(1) Cfr. C. Gini : *Di una estensione del concetto di scostamento medio e di alcune applicazioni alla misura dei caratteri qualitativi*, già cit.

(2) Cfr. C. Gini : *Di una estensione*, ecc., già cit., e cfr., anche quanto è detto in argomento alle pagg. 149-150.

(3) L'analogia alla quale ricorre l'A. nel tentativo di convincerci della necessità di adottare due definizioni diverse nei due casi, è solo formale, e, a parte il fatto che una economia nella quale si barattino esclusivamente due merci non esiste, non regge quando si paragoni l'ufficio della moneta negli scambi con quello della media aritmetica nella misura della disuguaglianza.

(4) Cfr. G. Pietra : *Note di Statistica Metodologica*, già cit., pag. 6 dell'estratto.

(5) Cfr. C. Gini : *Variabilità e Mutabilità*, già cit.

L'attitudine a variare di un carattere è espressa dalla disuguaglianza della sua distribuzione. Neppure la disuguaglianza tra i dati numerici che esprimono l'intensità del carattere nei varii casi è, così definita, una *quantità* suscettibile di misura. Sono quantità, invece, suscettibili di misura: la disuguaglianza tra due dati di posto determinato nella graduatoria delle intensità del carattere (campo di variazione, differenza interquartile, ecc.), la disuguaglianza media, di qualsiasi ordine, rispetto ad una media (scostamenti medi, di qualsiasi ordine, da una media); le disuguaglianze medie, di qualsiasi ordine, dei dati fra di loro (differenze medie, di qualsiasi ordine), ecc. In generale ogni funzione dei termini della serie che esprima disuguaglianza, comunque definita, tra questi, è una misura di disuguaglianza e rende, dal particolare punto di vista espresso dalla propria definizione, un aspetto della variabilità del carattere. E può dirsi per questo un indice di variabilità del carattere.

Anche la misura della concentrazione, che è un particolare aspetto della variabilità, è una misura di disuguaglianza e un indice di variabilità.

L'importanza ed il campo di applicazione di ciascun indice sono nettamente segnati dalle definizioni dalle quali essi discendono. L'uso di ciascuno di essi è pienamente legittimo quando a ciascuno si chiede ciò che esso può dire. Le loro discordanze non si contraddicono. La loro coesistenza lungi dall'essere fonte di confusione, aggiunge informazioni a informazioni. E se confusione ingenera in qualche Autore, ciò non può ascriversi certo a colpa degli indici o di chi tali indici costruisce.

PROF. CORRADO GINI, *Direttore responsabile.*

Roma — Tipografia Ippolito Failli